

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 08:28:29 ; Search time 34.13 Seconds
(without alignments)
27.760 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209
Sequence: 1 A1SYGNSADAQPYVAKIGVDAKQINGKNTAVGIYAGYN 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	25.4	1235	1 W21572	Rat brain serine e
2	53	25.4	1249	1 W21571	Rat brain homology
3	52	24.9	199	1 R34398	Helicobacter pylori
4	50	23.9	833	1 W27148	HMG-CoA reductase
5	49.5	23.7	255	1 W55201	H. pylori ORF 01cp
6	49.5	23.7	255	1 W55477	H. pylori ORF 07ep
7	49	23.4	540	1 R48667	Chitinase 1. New c
8	48.5	23.2	435	1 R32999	Rat choline kinase
9	48.5	23.2	1156	1 W46856	Bacillus thuringie
10	48	23.0	341	1 R70142	Porcine mutarotase
11	48	23.0	341	1 R72964	Pig kidney cell mu
12	48	23.0	532	1 R47237	Wild-type feline H
13	48	23.0	532	1 R36588	Feline herpesvirus
14	48	23.0	532	1 W60432	Feline herpesvirus
15	47	22.5	541	1 W44165	Endonuclease cauda
16	46.5	22.2	507	1 P50263	Sequence encoded b
17	46	22.0	165	1 W38713	S. pneumoniae glut
18	46	22.0	493	1 R48669	Chitinase derivati
19	46	22.0	778	1 W35004	Thermotoga maritim
20	46	22.0	832	1 W74089	Human HPT-1 protei
21	46	22.0	1287	1 R79944	Helicobacter pylori
22	46	22.0	1291	1 R75201	Tyrosine phosphata
23	46	22.0	1296	1 R41198	CT. Recombinant He
24	46	22.0	2893	1 W1356	Helicobacter polyp
25	46	22.0	2893	1 W98828	Minor ampullate sp
26	45.5	21.8	251	1 R60192	HAPPV p50 spindle
27	45.5	21.8	351	1 R52635	Streptomyces sp. t
28	45.5	21.8	407	1 W61249	Streptococcus pneu
29	45.5	21.8	592	1 W52304	Glucosyltransferase pr
30	45.5	21.8	1051	1 W52304	HPV-3 JS isolate
31	45.5	21.8	2233	1 W48711	HPV-3 JS isolate
32	45.5	21.8	2233	1 W48712	HPV-3 Vero CP45 v
33	45.5	21.8	2233	1 W48713	HPV-3 Vero CP45 v
34	45	21.5	288	1 W27720	H. pylori VacA pro

35	45	21.5	453	1 W17784	FIV integrase-LexA
36	45	21.5	505	1 R63807	Protein disulphide
37	45	21.5	514	1 W04346	Chamaecyparis obtu
38	45	21.5	514	1 W42122	Japanese cypress p
39	45	21.5	542	1 R48668	Chitinase 2. New c
40	45	21.5	547	1 R67385	Mitochondrial prot
41	45	21.5	668	1 W19668	Human alanyl-tRNA
42	45	21.5	1004	1 V07088	Renal cancer assoc
43	45	21.5	1004	1 V07088	FIV ppr clone 34 O
44	45	21.5	1124	1 W53668	H. pylori GHPO 374
45	45	21.5	1290	1 W38269	IgG-Fc binding pro
45	45	21.5	2594	1 W14748	

ALIGNMENTS

RESULT 1	
W21572	20-AUG-1997 (first entry)
ID	W21572: standard; Protein; 1235 AA.
AC	W21572:
DE	Rat brain serine ectopeptidase.
DE	Serine ectopeptidase; tripeptidylpeptidase II; TPP II;
KW	alternative splicing; rodent; rat; cerebral cortex;
KW	neuropeptide; cholecystokinin; CCK; inactivation; degradation;
KW	anorexia; schizophrenia; Parkinson's disease; depression;
KW	irritable bowel syndrome; bulimia; pathological obesity.
OS	Rattus sp.
FT	Key
FT	region
FT	85..1235
FT	/note- "The C-terminal region of the protein encoded
FT	by the alternatively spliced clone has a
FT	sequence identical to this region"
PN	W09635805-A2.
PN	14-NOV-1996.
PR	09-MAY-1996; F00700.
PR	09-MAY-1995; FR-005489.
PA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI	Bambard RB, Bishop PB, Bourgeat P, Chan S, Ganelin CR,
PI	Leblond B, Moore AMT, Schwartz JC, Vargas F;
PI	Lihua Z, Rose C;
DR	WPI: 96-518693/51.
DR	N-PSDB: T72654.
PT	Screening medicaments for treating disorders linked to inactivation
PT	of endogenous neuro-peptide(s) - by contacting candidate molecule
PT	with membrane tri:peptidyl-peptidase (homologue) and measuring
PS	enzyme activity
PS	Claim 4: Pages 166-171; 212pp; French.
CC	An enzyme with specificity for cholecystokinin (CCK) substrates
CC	(specifically the non-sulphated CCK8 and the CCK5 peptides) was
CC	purified from rat cerebral cortex membranes using high performance
CC	liquid chromatography. Based on amino acid sequence data from the
CC	purified enzyme, PCR primers were designed to amplify oligonucleotide
CC	probes A and B of lengths 350 and 380 nucleotides, respectively.
CC	Using these probes, two distinct clones were identified in a rat
CC	brain cDNA library. The first clone encoded the protein which is
CC	the rodent homologue of human tripeptidylpeptidase II. The second
CC	clone differed in the 5' region, probably as a result of alternative
CC	splicing, and encodes a protein of the present sequence which appears
CC	to be a serine ectopeptidase. The rat enzymes are preferred for use
CC	in a new method of screening for medicaments for treating disorders
CC	linked to the inactivation or degradation (or being treatable by
CC	retarding physiological degradation) of endogenous neuropeptides. In
CC	the method, a candidate molecule is contacted with a membrane
CC	tripeptidylpeptidase or homologue, and enzyme activity is measured.
CC	Disorders linked to inactivation or degradation of endogenous
CC	neuropeptides, include food intake disorders, cognitive and motor
CC	disorders such as anorexia, schizophrenia, Parkinson's disease and
CC	depression, as well as gastrointestinal transit disorders such as
CC	irritable bowel syndrome, bulimia and pathological obesity.
SO	Sequence 1235 AA;

CC regulate the degradation of HMG-CoA reductase, e.g. as therapeutic
CC agents to reduce hypercholesterolaemia, and to elucidate how the
CC cholesterol pathway modulates the degradation of HMG-CoA reductase. In
CC addition, as a result of their ability to bind the proteasome complex,
CC antibodies that specifically bind Hrd polypeptides can be used to
CC isolate the proteasome complex. Further, they can be used in various
CC assays to identify compounds that modify the degradation of HMG-CoA
CC reductase independently of the beneficial LDL receptor control axis.
CC The nucleic acid molecules can be used as molecular probes for the
CC isolation of homologous nucleic acid molecules and for the detection of
CC HRD nucleic acid molecules in yeast.
SQ Sequence 833 AA;

Query Match 23.9%; Score 50; DB 1; Length 833;
Best Local Similarity 37.8%; Pred. No. 39;
Matches 14; Conservative 7; Mismatches 10; Indels 6; Gaps 2;

OY 6 NSADAQPYV--GAKIGOVDAKQINGKNTAYGIYAGYN 40
DB 170 DSAKALLTYQRAAQIGNKAKOV---LAKYKYSGFN 202

RESULT 5
W55201
ID W55201 standard; Protein; 255 AA.

AC W55201.1998 (first entry)
DT 15-JUN-1998
DE H. pylori ORF 01cpl1414orf2 protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN W09737044-A1.

PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.

PA (ASTR) ASTRA AB.
PI Alm RA, Smith D.
PM WPI: 97-503122/46.
DR N-PSDB: V24610.

PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claim 14; Page 453; 1145pp; English.

CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or expected
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 255 AA;

Query Match 23.7%; Score 49.5; DB 1; Length 255;

Best Local Similarity 34.6%; Pred. No. 11;
Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5;

OY 3 SYG----NSADAQPYVGAKIGQVD-AKQIN-----GKNTAYGIY---AGYN 40
DB 101 TVGYSYNNANLS-FVGSKLGIMDGASQVNNFTYGVGFALNFEYSKRGYN 151

RESULT 6
W55477
ID W55477 standard; Protein; 255 AA.

AC W55477.
DT 30-JUN-1998 (first entry)
DE H. pylori ORF 07ep11916_5273452_c3_31 cell envelope OMP.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW bacterium.
OS Helicobacter pylori.
PN W09737044-A1.

PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.

PA (ASTR) ASTRA AB.
PI Alm RA, Smith D.
PM WPI: 97-503122/46.
DR N-PSDB: V24866.

PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 14,80; Pages 684-685; 1145pp; English.

CC This sequence is a H. pylori cell envelope outer membrane
CC protein (OMP) having a C-terminal tyrosine cluster motif
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The DNA and probes derived from it may be used for the
CC identification of H. pylori in a sample, and the diagnosis of
CC H. pylori infection. Nucleic acid sequences complementary to the
CC DNA act as antisense sequences, and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can
CC be used in immunoassays to evaluate the abundance and distribution
CC of H. pylori-specific antigens. The genomic sequence of H. pylori
CC (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were
CC analysed for ORF of at least 180 nucleotides, and the predicted
CC coding regions defined by computer evaluation. To identify likely
CC H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having
CC identified and determined the sequences of interest, particular
CC regions can be isolated from H. pylori by PCR amplification for
CC recombinant polypeptide production, e.g. in E. coli hosts.
SQ Sequence 255 AA;

Query Match 23.7%; Score 49.5; DB 1; Length 255;
Best Local Similarity 34.6%; Pred. No. 11;
Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5;

OY 3 SYG----NSADAQPYVGAKIGQVD-AKQIN-----GKNTAYGIY---AGYN 40
DB 101 TVGYSYNNANLS-FVGSKLGIMDGASQVNNFTYGVGFALNFEYSKRGYN 151

RESULT 7
R48667
ID R48667 standard; Protein; 540 AA.
AC R48667;

DT 12-OCT-1994 (first entry)
 DE Chitinase 1.
 KW Chitinase; derivative; beta-1,4-glycoside bond; chitin; yeast;
 KM beta-N-acetylhexosaminidase activity.
 OS Rhizopus oligosporus IF08631.
 FH Key
 FT peptide 1..22 Location/Qualifiers
 FT /note= "signal peptide"
 FT 23..540
 FT protein /note= "Mature protein"
 PN J06046849-A.
 PD 22-FEB-1994.
 PF 28-JUL-1992; 201427.
 PR 28-JUL-1992; JP-201427.
 PA (KAGO) KAGOME KK.
 DR WPI: 94-097015/12.
 DR N-PSDB: 056756.
 PT New chitinase enzyme and coding sequence - cleaves beta-1,4-glycoside bond of chitin but has no beta-N-acetylhexosaminidase activity
 PS Claim 1; Page 16-18; 66pp; Japanese.
 CC The sequences given in R48667-72 represent chitinase derivatives which cleave the beta-1,4-glycoside bond of chitin but have substantially no beta-N-acetylhexosaminidase activity. The DNA sequences encoding these proteins may be introduced into yeast and cultured for the production of the chitinase proteins.
 CC Sequence 540 AA;
 SQ

Query Match 23.4%; Score 49; DB 1; Length 540;
 Best Local Similarity 34.4%; Pred. No. 32;
 Matches 11; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

QY 5 GNSADAPYVYKIGVDYDLEGGASGYNAAF 36
 DB 145 GGSSDRPFQDAVIDGVLDLEGGASGYNAAF 176

RESULT 8
 ID R32999 standard; Protein; 435 AA.
 AC R32999;
 DT 14-JUN-1993 (first entry)
 DE Rat choline kinase.
 KW CK; epitope; tumour; diagnostic; antibodies; distribution; liver;
 KM cranial nerve disease.
 OS Rattus rattus.
 FH Key
 FT peptide 139..225 Location/Qualifiers
 FT peptide 278..327
 FT peptide 328..360
 FT peptide 398..412
 PN J05015367-A.
 PD 26-JAN-1993.
 PF 17-SEP-1991; 262516.
 PR 14-SEP-1990; JP-242528.
 PA (MITK) MITSUI TOATSU CHEM INC.
 DR WPI: 93-070176/09.
 DR N-PSDB: 037687.
 PT Rat choline kinase having specified DNA sequence - useful for diagnosing cranial nerve diseases, liver diseases, tumours, etc.
 PS Claim 1; Page 21; 26pp; Japanese.
 CC The choline kinase (CK) gene may be obt'd. from rat brain, human brain, human liver, chicken liver, egg, soy bean, etc. The rat CK gene is shown, and can be isolated by PCR using oligonucleotides or the amplified DNA fragments as primers. The oligonucleotides or the amplified DNA fragments, can then be used as probes to detect CK coding sequences. Rat CK may be purified by MMA-agarose affinity chromatography. A large amt. of CK may be produced and antibodies against CK may be obt'd. for use in detecting the distribution of CK in tissues or cells, and the diagnosis of cranial nerve disease, liver disease and tumours.
 CC Sequence 435 AA;
 SQ

Query Match 23.2%; Score 48.5; DB 1; Length 435;
 Best Local Similarity 34.4%; Pred. No. 29;
 Matches 11; Conservative 9; Mismatches 9; Indels 3; Gaps 1;

QY 2 ISYGNADAPYVYKAG---IGVDKAKOINKN 30
 DB 20 ISCGSAAPRPGVGOORDAAGELESKOLGCRS 51

RESULT 9
 ID W46856 standard; Protein; 1156 AA.
 AC W46856;
 DT 11-JUN-1998 (first entry)
 DE Bacillus thuringiensis toxin designated 86B1(a).
 KW Toxin; lepidopteran pest; control; Agrotis ipsilon; black cutworm;
 KM Heliothis virescens; Helicoverpa zea.
 OS Bacillus thuringiensis.
 PN W09800546-A2.
 PD 08-JAN-1998.
 PF 01-JUL-1997; U11658.
 PR 01-JUL-1996; US-674002.
 PA (MYCO) MYCOGEN CORP.
 PT Marva KE; Schnepf HE; Stockhof BA; Walz M; Wicker C;
 DR WPI: 98-086971/08.
 DR N-PSDB: V16515.
 PT New isolated Bacillus thuringiensis isolate(s) - used to obtain genes encoding toxins which are active against lepidopteran pests such as the Black cutworm
 PS Claim 34; Pages 82-86; 183pp; English.
 CC The present sequence represents a Bacillus thuringiensis toxin designated 86B1(a) which is active against lepidopteran pests. The toxin isolates can be used for the control of lepidopteran pests such as Agrotis ipsilon (black cutworm), Heliothis virescens and Helicoverpa zea. PCR primers and probes can be derived from the polynucleotide encoding the toxin and used for the amplification and detection of other CC toxin-encoding sequences.
 CC Sequence 1156 AA;
 SQ

Query Match 23.2%; Score 48.5; DB 1; Length 1156;
 Best Local Similarity 25.5%; Pred. No. 98;
 Matches 14; Conservative 8; Mismatches 16; Indels 17; Gaps 1;

QY 2 ISYGNADAPYVYKAGVDYDLEGGASGYNAAF 39
 DB 389 ISCGHTTATOTILGRNIFRVDSQCNLNDTTYGVNRAVYTHDASEGSRVYEGY 443

RESULT 10
 ID R70142 standard; Protein; 341 AA.
 AC R70142;
 DT 10-NOV-1995 (first entry)
 DE Porcine mutarotase (MUT) enzyme.
 KW Porcine mutarotase; MUT; blood glucose levels;
 KM recombinant production.
 OS Sus scrofa.
 PN J07039380-A.
 PD 10-FEB-1995.
 PF 30-JUL-1993; 208328.
 PR 30-JUL-1993; JP-208328.
 PA (AMANO) AMANO PHARM KK.
 DR WPI: 95-117860/16.
 DR N-PSDB: 083214.
 PT DNA encoding mutarotase and recombinant constructs comprising this - for the large scale production of mutarotase for its use in the determination of blood glucose levels
 PS Claim 3; Pages 7-8; 8pp; Japanese.
 CC 083214 encodes R70142 porcine kidney mutarotase (MUT). The DNA as part of a claimed expression vector can be used for the recombinant prodn. of the MUT enzyme. The MUT enzyme can be used for the

DT 15-FEB-1999 (first entry)
DE Feline herpesvirus glycoprotein E.
KW FHV: vaccine: immunisation: feline rhinotracheitis: vector.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 10:36:35 ; Search time 27.94 Seconds

(without alignments)
20.707 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209
Sequence: 1 A1SYGNSADNAPRYVGAKICQVDAKQINCKNTAYGTYAGYN 40Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0
Maximum DB seq length: 1000000Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	24.9	199	2	US-08-211-312-6
2	52	24.9	199	3	US-08-472-285-6
3	48	23.0	532	3	US-08-911-321-11
4	48	23.0	532	3	US-08-911-321-11
5	48	23.0	532	4	PCT-US95-13975-2
6	47.5	22.7	283	2	US-08-852-401-4
7	46	22.0	832	1	US-08-431-560-1
8	46	22.0	832	1	US-08-463-345-1
9	46	22.0	1287	1	US-08-200-232-2
10	46	22.0	1287	4	PCT-US95-02219-2
11	46	22.0	1287	4	PCT-US95-02219A-2
12	45.5	21.8	251	1	US-08-209-747-8
13	45.5	21.8	251	1	US-08-458-298-8
14	45.5	21.8	331	1	US-08-356-180-3
15	45.5	21.8	2233	2	US-08-569-853-1
16	45.5	21.8	2233	2	US-08-569-853-2
17	45	21.5	462	2	US-08-865-597A-2
18	45	21.5	626	2	US-08-956-242-2
19	45	21.5	658	1	US-08-426-236-2
20	44.5	21.3	359	1	US-08-457-997B-2
21	44	21.1	32	1	US-08-462-221-4
22	44	21.1	32	1	US-08-446-692-21
23	44	21.1	32	2	US-08-488-351A-21
24	44	21.1	74	2	US-08-343-443B-11
25	44	21.1	198	2	US-08-943-915-33
26	44	21.1	207	2	US-08-943-915-2
27	44	21.1	207	2	US-08-943-915-5
28	44	21.1	263	2	US-08-752-844-66
29	44	21.1	339	1	US-08-266-451B-22

30	44	21.1	339	2	US-08-748-725-22	Sequence 22, Appl
31	44	21.1	340	2	US-08-355-844-1	Sequence 1, Appl
32	44	21.1	340	4	PCT-US95-16126-1	Sequence 1, Appl
33	43.5	20.8	380	2	US-08-472-659-34	Sequence 34, Appl
34	43.5	20.8	380	2	US-08-474-661-34	Sequence 34, Appl
35	43.5	20.8	380	2	US-08-611-977-34	Sequence 34, Appl
36	43.5	20.8	449	4	PCT-US94-02539-2	Sequence 2, Appl
37	43.5	20.8	449	4	PCT-US94-02539-2	Sequence 2, Appl
38	43.5	20.8	459	2	US-08-673-312-2	Sequence 2, Appl
39	43.5	20.8	471	2	US-08-657-392-31	Sequence 31, Appl
40	43.5	20.8	471	2	PCT-US94-02539-31	Sequence 31, Appl
41	43.5	20.8	547	2	US-08-467-822-35	Sequence 35, Appl
42	43.5	20.8	573	2	US-08-706-209-1	Sequence 1, Appl
43	43.5	20.8	573	4	PCT-US94-06362-1	Sequence 1, Appl
44	43.5	20.8	573	4	PCT-US96-11373-1	Sequence 1, Appl
45	43.5	20.8	573	4	PCT-US96-11375-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-08-211-312-6
: Sequence 6, Application US/08211312
: Patent No. 5986051
:
: GENERAL INFORMATION:
: APPLICANT: LABIGNE, AGNES
: APPLICANT: CUSSAC, VALERIE
: APPLICANT: FERRERO, RICHARD
: TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY
: TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
: ADDRESS: P.C.
: STREET: 1755 S. Jefferson Davis Highway, Suite 400
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/211.312
: FILING DATE: 01-JUL-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 12198
: FILING DATE: 03-OCT-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR92/00921
: FILING DATE: 02-OCT-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Oblon, No. 5986051man F.
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 660-075-0XPCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 413-3000
: TELEFAX: (703) 413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 199 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-211-312-6
:
: Query Match 24.9%, Score 52; DB 2; Length 199;
```


ATTORNEY/AGENT INFORMATION:
NAME: Mueller, Lisa L.
REGISTRATION NUMBER: 38, 978
REFERENCE/DOCKET NUMBER: FER2159POO30US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5400
TELEFAX: 312-616-5460
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-852-401-4

Query Match 22.7%; Score 47.5; DB 2; Length 283;
Best Local Similarity 27.5%; Pred. No. 14;
Matches 11; Conservative 5; Mismatches 19; Indels 5; Gaps 1;

OY 6 NSADAOPYV-----GAKIGVDAKQINGKNTAYGIYAGYN 40
DB 172 NMGSGNPTPTTEVCAGAGSYIIDQALGAGCTTLLILNASN 211

RESULT 7
US-08-431-560-1
Sequence 1, Application US/08431560
Patent No. 5620855
GENERAL INFORMATION:
APPLICANT: Anne H. Dantzig, et al.
TITLE OF INVENTION: Mammalian Influx Peptide
TITLE OF INVENTION: Transporter
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: U.S.A.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,560
FILING DATE: 01-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/013,462
FILING DATE: 04-FEB-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-431-560-1

Query Match 22.0%; Score 46; DB 1; Length 832;
Best Local Similarity 36.4%; Pred. No. 95;
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 14 VGAKIGVDAKQINGKNTAYGI 35
DB 581 IGTKVGNTAKDEGLDISYSL 602

RESULT 8

US-08-463-345-1
Sequence 1, Application US/08463345
Patent No. 5710018
GENERAL INFORMATION:
APPLICANT: Anne H. Dantzig, et al.
TITLE OF INVENTION: Mammalian Influx Peptide
TITLE OF INVENTION: Transporter
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: U.S.A.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,345
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/013,462
FILING DATE: 04-FEB-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-345-1

Query Match 22.0%; Score 46; DB 1; Length 832;
Best Local Similarity 36.4%; Pred. No. 95;
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 14 VGAKIGVDAKQINGKNTAYGI 35
DB 581 IGTKVGNTAKDEGLDISYSL 602

RESULT 9
US-08-200-232-2
Sequence 2, Application US/08200232
Patent No. 5721349
GENERAL INFORMATION:
APPLICANT: Cover, Timothy L.
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
TITLE OF INVENTION: AND RELATED METHODS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,232
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spratl, Gwendolyn D.

REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-200-232-2

Query Match 22.0%; Score 46; DB 1; Length 1287;
Best Local Similarity 40.5%; Pred. No. 1.6e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADAPYVGAKIGVDAKINGKNTAYGIYAGYN 40
Db 1040 YGTSAGVDAYLN--GQVEA--IVGGFGSYG-YSSFN 1070

RESULT 10
PCT-US95-02219-2
Sequence 2, Application PC/TUS9502219
GENERAL INFORMATION:
APPLICANT: Cover, Timothy L.
APPLICANT: Blaser, Martin J.
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
TITLE OF INVENTION: AND RELATED METHODS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02219
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02219-2

Query Match 22.0%; Score 46; DB 4; Length 1287;
Best Local Similarity 40.5%; Pred. No. 1.6e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADAPYVGAKIGVDAKINGKNTAYGIYAGYN 40
Db 1040 YGTSAGVDAYLN--GQVEA--IVGGFGSYG-YSSFN 1070

RESULT 11
PCT-US95-02219A-2
Sequence 2, Application PC/TUS9502219A
GENERAL INFORMATION:
APPLICANT: Cover, Timothy L.
APPLICANT: Tummuru, Murali KR
APPLICANT: Cao, Ping
APPLICANT: Thompson, Stuart A.
APPLICANT: Blaser, Martin J.
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
TITLE OF INVENTION: AND THE RELATED METHODS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02219A
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02219A-2

Query Match 22.0%; Score 46; DB 4; Length 1287;
Best Local Similarity 40.5%; Pred. No. 1.6e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADAPYVGAKIGVDAKINGKNTAYGIYAGYN 40
Db 1040 YGTSAGVDAYLN--GQVEA--IVGGFGSYG-YSSFN 1070

RESULT 12
US-08-209-747-8
Sequence 8, Application US/08209747
Patent No. 5733771
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Colgin, Mark
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
TITLE OF INVENTION: Silk Proteins
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

Tue May 23 16:14:18 2000

us-09-164-714-1.rai

Page 7

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	21.8%	Score 45.5	DB 2	length 2233
Best Local Similarity	35.3%	Pred. No. 3,9e+02		
Matches	12	Conservative	6	Mismatches 15; Indels 1; Gaps 1;

OY	1	AISTGNSAD	~QPYVGARIGQVDARQINGKNTAY	33
Db	438	AISTENADVQSFQIGLIFENKFI	EPQDLEDLLTY	471

Search completed: May 20, 2000, 12:18:21
Job time: 6106 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 11:10:21 ; Search time 43.2 Seconds

(without alignments)
54.287 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209

Sequence: 1 AISTGNSADAPYVCAKIGQVDAKQINGKNTAYGIYAGYN 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR.63:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	59.5	28.5	380	T24786	hypothetical prote
2	55.5	26.6	289	S38806	porin opma - Rhodo
3	55.5	26.6	1196	S40908	TYB protein - yeas
4	55.5	26.6	1278	B22671	TYB protein - yeas
5	55.5	26.6	1328	S52481	TYB protein - yeas
6	55.5	26.6	1328	S52894	TYB protein - yeas
7	55.5	26.6	1328	S50948	TYB protein - yeas
8	55.5	26.6	1328	B28097	TYB protein - yeas
9	55.5	26.6	1348	B23496	TYB protein - yeas
10	55.5	26.6	1749	S50972	TYB protein - yeas
11	55.5	26.6	1755	S50641	TYB protein YER138
12	55.5	26.6	1755	S50663	TYB protein - yeas
13	55.5	26.6	1755	S57047	TYB protein - yeas
14	55.5	26.6	1755	S57045	TYB protein - yeas
15	55.5	26.6	1755	S45736	TYB protein - yeas
16	55.5	26.6	1755	S69963	TYB protein - yeas
17	55.5	26.6	1755	S69839	TYB protein - yeas
18	55.5	26.6	1755	S69980	TYB protein - yeas
19	55.5	26.6	1755	S69957	TYB protein - yeas
20	55.5	26.6	1755	S69951	TYB protein - yeas
21	55.5	26.6	1755	S69969	TYB protein - yeas
22	55.5	26.6	1755	S69979	TYB protein - yeas
23	55.5	26.6	1755	S69979	TYB protein - yeas
24	55.5	26.6	1755	S69838	TYB protein - yeas
25	55.5	26.6	1755	S70238	TYB protein - yeas
26	55.5	26.6	1755	S69955	TYB protein - yeas
27	55.5	26.6	1755	S69845	TYB protein - yeas
28	55.5	26.6	1755	S69866	TYB protein - yeas
29	55.5	26.6	1755	S69982	TYB protein - yeas
30	55.5	26.6	1755	S69949	TYB protein - yeas

31	55.5	26.6	1755	2	S61763	TYB protein - yeas
32	55.5	26.6	1755	2	S69975	TYB protein - yeas
33	55.5	26.6	1756	2	S43867	TYB protein - yeas
34	55.5	26.6	1756	2	S69963	TYB protein - yeas
35	55.5	26.6	1770	2	S45842	TYB protein - yeas
36	55.5	26.6	1770	2	S69953	TYB protein - yeas
37	55.5	26.6	1770	2	S69948	TYB protein - yeas
38	55.5	26.6	1770	2	S70233	TYB protein - yeas
39	55.5	26.6	1770	2	S70230	TYB protein - yeas
40	55.5	26.6	1770	2	S69966	TYB protein - yeas
41	55.5	26.6	1770	2	S69950	TYB protein - yeas
42	55.5	26.6	1770	2	S58651	TYB protein - yeas
43	55.5	26.6	1771	2	S53592	TYB protein - yeas
44	55.5	26.6	1793	2	S52601	TYB protein - yeas
45	55.5	26.6	1810	2	S69973	TYB protein - yeas

ALIGNMENTS

```
RESULT 1
T24786
hypothetical protein ZK1321.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T24786; T27760
R:Gardner, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z19935
A:Accession: T24786
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-380 <WIL>
A:Cross-references: EMBL:Z48717; PIDN:CAA88612.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone T10B9
R:Gardner, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z20415
A:Accession: T27760
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-380 <WT2>
A:Cross-references: EMBL:Z48584; PIDN:CAA88478.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone ZK1321
C:Genetics:
A:Gene: CESP:ZK1321.4
A:Map position: 2
A:Introns: 4/3; 35/2; 85/3; 115/2; 224/2; 328/2; 346/3

Query Match      28.5%; Score 59.5; DB 2; Length 380;
Best Local Similarity 38.6%; Pred. No. 2.7;
Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

OY      3 SYGNSADAPYVCAKIGQVDAKQIN-----GNTAYGIYAG 38
DB      224 SYNGNATSQPY-GAGSGCTPLNQMTFINTSPAFCANGAYGACCG 266

RESULT 2
S38806
porin opma - Rhodospseudomonas blastica
C:Species: Rhodospseudomonas blastica
C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 07-May-1999
C:Accession: S38806; S38793
R:Kreusch, A.; Neubueser, A.; Schiltz, E.; Weckesser, J.; Schulz, G.E.
Protein Sci. 3: 58-63, 1994
A:Title: Structure of the membrane channel porin from Rhodospseudomonas blastica at 2.
A:Reference number: S38806; MUID:94191532
A:Accession: S38806
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-289 <KRE>
```

R.Kreusch, A.; Neubueser, A.; Schiltz, E.; Weckesser, J.; Schulz, G.E.
submitted to the Protein Sequence Database, November 1993

A:Reference number: S38793

A:Accession: S38793

A:Molecule type: protein

A:Residues: 1-71,'X',73-82;91-104;109-137;146-163;173-289 <KR2>

C:Genetics:

A:Gene: opma

C:Keywords: homotrimer; membrane protein

Query Match 26.6%; Score 55.5; DB 2; Length 289;
Best Local Similarity 41.7%; Pred. No. 6.8;

Matches 15; Conservative 3; Mismatches 15; Indels 3; Gaps 1;

Oy 4 YGNSADAPYVGAKIGQV--DAKOINGKNT 39

Db 227 YGNVAFGATTVRAVYSDIDRA--GADTANGIGADY 259

RESULT 3
S40908
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.A

N:Alternate names: protein YAR009c

C:Species: Saccharomyces cerevisiae

C>Date: 13-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 22-Oct-1999

C:Accession: S40908

R:Clark, M.W.; Keng, T.; Storms, R.R.; Zhong, W.; Fortlin, N.; Zeng, B.; Delaney, S.; Que
submitted to the EMBL Data Library, November 1993

A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: Analysis of the 4

A:Reference number: S40891

A:Accession: S40908

A:Molecule type: DNA

A:Residues: 1-1196 <CLA>

A:Cross-references: EMBL:L22015; NID:g1339990; PIDN:AAC04967.1; PID:g2564964

C:Genetics:

A:Map position: 1R

A:Mobile element: retrotransposon Ty1.A

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1196;
Best Local Similarity 43.8%; Pred. No. 28;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKOINGKNT 31

Db 1053 SYGN---QPYKSOIGNITLNGKVIIGKST 1080

RESULT 4
B2671
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty912

C:Species: Saccharomyces cerevisiae

C>Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 26-Aug-1999

C:Accession: B2671; S05890

R:Clare, J.; Farabaugh, P.

Proc. Natl. Acad. Sci. U.S.A. 82, 2829-2833, 1985

A:Title: Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of

A:Reference number: A94039; MUID:85190589

A:Accession: B2671

A:Molecule type: DNA

A:Residues: 1-1278 <CLA>

R:Mellor, J.; Fulton, S.M.; Dobson, M.J.; Wilson, W.; Kingsman, S.M.; Kingsman, A.J.

Nature 313, 243-246, 1985

A:Title: A retrovirus-like strategy for expression of a fusion protein encoded by yeast

A:Reference number: S05890

A:Accession: S05890

A:Molecule type: DNA

A:Residues: 1-6,'I','8-10','K',12-53 <MEU>

A:Cross-references: EMBL:X01736; NID:g4708; PIDN:CAA25874.1; PID:g1326020

C:Genetics:

A:Mobile element: retrotransposon Ty912

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1278;
Best Local Similarity 43.8%; Pred. No. 30;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKOINGKNT 31

Db 1135 SYGN---QPYKSOIGNITLNGKVIIGKST 1162

RESULT 5
S52481
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1

N:Alternate names: protein YML039w

C:Species: Saccharomyces cerevisiae

C>Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999

C:Accession: S52481

R:Connor, R.; Churchev, C.

submitted to the EMBL Data Library, February 1995

A:Reference number: S52478

A:Accession: S52481

A:Molecule type: DNA

A:Residues: 1-1328 <CON>

A:Cross-references: EMBL:248430; NID:g683664; PIDN:CAA88330.1; PID:e155222; PID:g1326

C:Genetics:

A:Map position: 13L

A:Mobile element: retrotransposon Ty1

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 32;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKOINGKNT 31

Db 1185 SYGN---QPYKSOIGNITLNGKVIIGKST 1212

RESULT 6
S52894
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-9532

N:Alternate names: protein YM9532.10c; protein YMR045c

C:Species: Saccharomyces cerevisiae

C>Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999

C:Accession: S52894

R:Odell, C.; Bowman, S.

submitted to the EMBL Data Library, February 1995

A:Reference number: S52885

A:Accession: S52894

A:Molecule type: DNA

A:Residues: 1-1328 <ODE>

A:Cross-references: EMBL:248502; NID:g695715; PIDN:CAA88411.1; PID:e159526; PID:g1326

C:Genetics:

A:Map position: 13R

A:Mobile element: retrotransposon Ty1-9532

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 32;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKOINGKNT 31

Db 1185 SYGN---QPYKSOIGNITLNGKVIIGKST 1212

RESULT 7
S50948

Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1

N:Alternate names: protein YM9827.08; protein YML045W
 C:Species: Saccharomyces cerevisiae
 C:Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 05-Nov-1999
 C:Accession: S50948
 R:Odell, C.; Bowman, S.
 submitted to the EMBL Data Library, January 1995
 A:Reference number: S50941
 A:Accession: S50948
 A:Molecule type: DNA
 A:Residues: 1-1328 <ODE>
 A:Cross-references: EMBL:Z47816; NID:9642303; PIDN:CAA87830.1; PID:e135608; PID:g1326015
 C:Genetics:
 A:Map position: 13L
 A:Mobile element: retrotransposon Ty1
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;
 Best Local Similarity 43.8%; Pred. No. 32;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
 |||| ||| :|| : |||||
 Db 1185 SYGN---QPYKKSQIGNIYLNGKVIKGKST 1212

RESULT 8
 B28097
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-H3 (Fragment)
 C:Species: Saccharomyces cerevisiae
 C:Date: 03-Nov-1988 #sequence_revision 25-Apr-1997 #text_change 25-Apr-1997
 C:Accession: B28097
 R:Boeke, J.D.; Elchinger, D.; Castillon, D.; Fink, G.R.
 Mol. Cell. Biol. 8, 1432-1442, 1988
 A:Title: The Saccharomyces cerevisiae genome contains functional and nonfunctional copies
 A:Reference number: A28097; MUID:88246410
 A:Accession: B28097
 A:Molecule type: DNA
 A:Residues: 1-1328 <BOE>
 A:Cross-references: EMBL:M18706
 C:Genetics:
 A:Mobile element: retrotransposon Ty1-H3
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;
 Best Local Similarity 43.8%; Pred. No. 32;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
 |||| ||| :|| : |||||
 Db 1185 SYGN---QPYKKSQIGNIYLNGKVIKGKST 1212

RESULT 9
 B23496
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-17
 C:Species: Saccharomyces cerevisiae
 C:Date: 20-Aug-1987 #sequence_revision 20-Aug-1987 #text_change 25-Apr-1997
 C:Accession: B23496
 R:Warmington, J.R.; Waring, R.B.; Newton, C.S.; Indge, K.J.; Oliver, S.G.
 Nucleic Acids Res. 13, 6679-6693, 1985
 A:Title: Nucleotide sequence characterization of Ty 1-17, a class II transposon from yeast
 A:Reference number: A93591; MUID:86041864
 A:Accession: B23496
 A:Molecule type: DNA
 A:Residues: 1-1348 <MAR>
 A:Cross-references: EMBL:X03840; EMBL:J01333; EMBL:K00633; EMBL:X02991
 C:Genetics:
 A:Mobile element: retrotransposon Ty1-17
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1348;
 Best Local Similarity 43.8%; Pred. No. 32;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
 |||| ||| :|| : |||||
 Db 1205 SYGN---QPYKKSQIGNIYLNGKVIKGKST 1232

RESULT 10
 S69972
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.N
 N:Alternate names: protein N2453
 C:Species: Saccharomyces cerevisiae
 C:Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 26-Aug-1999
 C:Accession: S69972
 R:Berges, P.; Dolignon, F.; Crouzet, M.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S62975
 A:Accession: S69972
 A:Molecule type: DNA
 A:Residues: 1-1749 <BER>
 A:Cross-references: EMBL:Z71330; NID:q1301918; PIDN:CAA95924.1; PID:e239892; PID:q130
 A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu
 C:Genetics:
 A:Map position: 14L
 A:Mobile element: retrotransposon Ty1.N
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1749;
 Best Local Similarity 43.8%; Pred. No. 42;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
 |||| ||| :|| : |||||
 Db 1606 SYGN---QPYKKSQIGNIYLNGKVIKGKST 1633

RESULT 11
 S50641
 TYB protein YER138C - yeast (Saccharomyces cerevisiae) retrotransposon Ty
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 26-Aug-1999
 C:Accession: S50641
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lamb
 A:Reference number: S50428
 A:Accession: S50641
 A:Molecule type: DNA
 A:Residues: 1-1755 <DIE>
 A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB6465.1; PID:g603378; MIPS:YER1
 C:Genetics:
 A:Map position: 5R
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;
 Best Local Similarity 43.8%; Pred. No. 42;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
 |||| ||| :|| : |||||
 Db 1612 SYGN---QPYKKSQIGNIYLNGKVIKGKST 1639

RESULT 12
 S50663
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.E
 N:Alternate names: protein YER160C
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 26-Aug-1999

C:Accession: S50663; S30812; S53556
R:Dietch, F.S.
Submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmids 8229, 9115, 9132, 9981, and lambda
A:Reference number: S50428
A:Accession: S50663
A:Molecule type: DNA
A:Residues: 1-1755 <DIE>
A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64687.1; PID:g603400
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,
submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30812
A:Molecule type: DNA
A:Residues: 1-230 <MUL>
A:Cross-references: EMBL:L10718
C:Genetics:
A:Map position: 5R
A:Mobile element: retrotransposon Ty1.E
C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 42;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGOV---DAKOINGKNT 31
Db 1612 SYGN---QPYKKSQIGNIYLNGKVIYGKST 1639

RESULT 13
S57047
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N:Alternate names: protein j1570; protein YJR029W
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herbe
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57040
A:Accession: S57047
A:Molecule type: DNA
A:Residues: 1-1755 <HUA>
A:Cross-references: EMBL:249528
A:Note: Biosynthesis of this protein involves a +1 frameshift in the codon for residue 4
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herbe
Yeast 11, 1179-1186, 1995
A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading fr
A:Reference number: S60503; MUID:96109930
A:Accession: S60512
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'TISTTTF', 435-1755 <ZAG>
A:Cross-references: EMBL:X87297
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Note: the difference at the amino end is due to translation from an incorrect reading
C:Genetics:
A:Map position: 10R
A:Mobile element: retrotransposon Ty1.JR
C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 42;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGOV---DAKOINGKNT 31
Db 1612 SYGN---QPYKKSQIGNIYLNGKVIYGKST 1639

RESULT 14
"

S57045
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N:Alternate names: protein j1560; protein YJR027W
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
A:Reference number: S57045
A:Accession: S57045
A:Molecule type: DNA
A:Residues: 1-1755 <HUA>
A:Cross-references: EMBL:249526
A:Note: Biosynthesis of this protein involves a +1 frameshift in the codon for residu
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; He
Yeast 11, 1179-1186, 1995
A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading
A:Reference number: S60503; MUID:96109930
A:Accession: S60511
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'TISTTTF', 435-1755 <ZAG>
A:Cross-references: EMBL:X87297
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Note: the difference at the amino end is due to translation from an incorrect read
C:Genetics:
A:Map position: 10R
A:Mobile element: retrotransposon Ty1.JR
C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 42;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGOV---DAKOINGKNT 31
Db 1612 SYGN---QPYKKSQIGNIYLNGKVIYGKST 1639

RESULT 15
S45736
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1
N:Alternate names: protein YBL004w-a; protein YBL0325
C:Species: Saccharomyces cerevisiae
C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Nov-1997
C:Accession: S45736; S45735
R:Delaveau, T.; Jacq, C.; Perea, J.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45736
A:Accession: S45736
A:Molecule type: DNA
A:Residues: 1-628
A:Cross-references: EMBL:Z35765; MIPS:YBL004w-a
A:Experimental source: strain S288C
R:Rieger, M.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45734
A:Accession: S45735
A:Molecule type: DNA
A:Residues: 617-1755 <RIE>
A:Cross-references: EMBL:Z35765; MIPS:YBL004w-a
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 2L
A:Mobile element: retrotransposon Ty1
C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 42;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADADOPYVYGAKEGOV---DAKOLNGKNT 31
||| :|| : |||||
Db 1612 SYGN---QPYKSOIGNITLNGKVIGKST 1639

Search completed: May 20, 2000, 12:19:13
Job time: 4132 sec

THIS PAGE BLANK (USPTO)


```
RESULT 2
YMO9_YEAST STANDARD; PRT; 1328 AA.
AC 003434;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YML039W OR YM8054.04.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Connor R., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: Z48430; CAA88330.1; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 151036 MW; C7D14E1AA675E93 CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAQPYVAGAKIGQV---DAKQINGKNT 31
Db 1185 SYGN---QPYKRSQIGNITLNGKVIYGKST 1212

RESULT 3
YME4_YEAST STANDARD; PRT; 1328 AA.
AC 004711;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YML044W OR YM8827.08.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
```

```
DR EMBL: Z47816; CAA87830.1; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 150945 MW; E9B964CD76C3281F CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAQPYVAGAKIGQV---DAKQINGKNT 31
Db 1185 SYGN---QPYKRSQIGNITLNGKVIYGKST 1212

RESULT 4
YMT5_YEAST STANDARD; PRT; 1328 AA.
AC 004214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YMR045C OR YM9532.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: Z48502; CAA88411.1; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 151187 MW; AA19E50B62B43F95 CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAQPYVAGAKIGQV---DAKQINGKNT 31
Db 1185 SYGN---QPYKRSQIGNITLNGKVIYGKST 1212

RESULT 5
YMO0_YEAST STANDARD; PRT; 1328 AA.
AC 004670;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YMR050C OR YM9796.03C.
OS Saccharomyces cerevisiae (Baker's yeast).
```



```
CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z49703; CAAB9760.1; -
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 150967 MW; DACD7A471697DIDD CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGARIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSOIGNIFLNGKVI GSKST 1212

RESULT 6
YCB9_YEAST STANDARD: PRT; 1347 AA.
AC P25384;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSPOSON TY1-17 PROTEIN B.
GN TY1B OR YCL019W OR YCL19W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86205247.
RA Warrington J.R., Anwar R., Newton C.S., Waring R.B., Davies R.W.,
RA Indge K.J., Oliver S.G.;
RT "A 'hot-spot' for Ty transposition on the left arm of yeast
RT chromosome III";
RL Nucleic Acids Res. 14:3475-3485(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86041864.
RA Warrington J.R., Waring R.B., Newton C.S., Indge K.J., Oliver S.G.;
RT "Nucleotide sequence characterization of Ty 1-17, a class II
RT transposon from yeast.";
RL Nucleic Acids Res. 13:6679-6693(1985).
RN [3]
RP SEQUENCE FROM N.A.
RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA Steve L.I.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X03840; CAA27458.1; -
DR EMBL; X59720; E264443; -
DR PIR; B23496; B23496.
DR PIR; S19345; S19345.
KW Transposable element; Hypothetical protein; Hydrolase;
FT ACT_SITE 1223 1230 ATP (POTENTIAL).
FT NP_BIND 1223 1230
SQ SEQUENCE 1347 AA; 154069 MW; AD3660C5E7B282FF CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1347;
Best Local Similarity 43.8%; Pred. No. 18;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGARIGOV---DAKQINGKNT 31
Db 1204 SYGN---QPYKSOIGNIFLNGKVI GSKST 1231

RESULT 7
YJ27_YEAST STANDARD: PRT; 1755 AA.
AC P47098; P87194;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSPOSON TY1 PROTEIN B.
GN TY1B OR YUR027W OR J1560.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96109930.
RA Zagulski M., Babinska B., Gromadka R., Migdalski A., Rytko J.,
RA Sulicka J., Herbert C.J.;
RT "The sequence of 24.3 kb from chromosome X reveals five complete open
RT reading frames, all of which correspond to new genes, and a tandem
RT insertion of a Ty1 transposon.";
RL Yeast 11:1179-1186(1995).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z49526; CAA89553.1; -
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFM; PF01021; TYA; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KW Aspartyl protease; ATP-binding.
FT ACT_SITE 461 461 PROTEASE (BY SIMILARITY).
FT NP_BIND 1631 1638 ATP (POTENTIAL).
SQ SEQUENCE 1755 AA; 198615 MW; 04E248A77FD7596F CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1755;
Best Local Similarity 43.8%; Pred. No. 23;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGARIGOV---DAKQINGKNT 31
Db 1612 SYGN---QPYKSOIGNIFLNGKVI GSKST 1639
```

```

RESULT 8
Y129_YEAST STANDARD; PRT; 1755 AA.
ID Y129_YEAST
AC P47100; P87195;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DE TRANSPOSON TY1 PROTEIN B.
GN TY1B OR YJR029W OR J1570.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96109930.
RA Zagulski M., Babinska B., Gromadka R., Migdalski A., Rylka J.,
RA Sulicka J., Herbert C.J.;
RT "The sequence of 24.3 kb from chromosome X reveals five complete open
RT reading frames, all of which correspond to new genes, and a tandem
RT insertion of a Ty1 transposon."
RL Yeast 11:1179-1186(1995).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY 023.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z49528; CAA8556.1; -
CC DR PROSITE; PS00141; ASP_PROTEASE; 1.
CC DR PFAM; PF01021; TYA; 1.
CC KM Transposable element; Hypothetical protein; Hydrolase;
CC Aspartyl protease; ATP-binding.
CC FT ACT SITE 461 461 PROTEASE (BY SIMILARITY).
CC FT NF_BIND 1631 1638 ATP (POTENTIAL).
CC SQ SEQUENCE 1755 AA; 198592 MW; C8B125BF2967C990 CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1755;
Best Local Similarity 43.8%; Pred. No. 23;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGOV---DAKQINGKNT 31
Db 1612 SYGN----QPYKSOQINIVILNGKVIAGKST 1639

RESULT 9
DPOD_DROME STANDARD; PRT; 1092 AA.
ID DPOD_DROME
AC P54358;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE DELTA CATALYTIC CHAIN (EC 2.7.7.7).
GN POLD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96125196.
RA Chiang C.S., Lehman I.R.;
RT "Isolation and sequence determination of the cDNA encoding DNA
RT polymerase delta from Drosophila melanogaster."
RL Gene 166:237-242(1995).
CC -1- FUNCTION: POSSSESSES TWO ENZYMATIC ACTIVITIES: DNA SYNTHESIS
CC (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT DEGRADERS SINGLE
CC STRANDED DNA IN THE 3' TO 5' DIRECTION. REQUIRED WITH ITS

```

```

CC ACCESSORY PROTEINS (PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) AND
CC REPLICATION FACTOR C (RFC) OR ACTIVATOR 1) FOR LEADING STRAND
CC SYNTHESIS. ALSO INVOLVED IN COMPLETING OKAZAKI FRAGMENTS INITIATED
CC BY THE DNA POLYMERASE ALPHA/PRIMASE COMPLEX (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE ->
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: HETERODIMER WITH SUBUNITS OF 125 KD AND 50 KD. THE 125 KD
CC SUBUNIT CONTAINS THE POLYMERASE ACTIVE SITE AND MOST LIKELY THE
CC ACTIVE SITE FOR THE 3'-5' EXONUCLEASE ACTIVITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X88928; CAA61369.1; -
CC DR FLYBASE; FBgn0012066; DNAPol-delta.
CC DR PRINTS; PR00106; DNAPOLB.
CC DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
CC DR PFAM; PF00136; DNA_pol_B; 1.
CC KM Transferrase; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Hydrolase; Exonuclease; Zinc-finger; Nuclear protein.
CC FT DOMAIN 4 19 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT ZN_FING 997 1017 C4-TYPE (POTENTIAL).
CC FT ZN_FING 1046 1064 C4-TYPE (POTENTIAL).
CC SQ SEQUENCE 1092 AA; 124879 MW; 976B640BB37DBD CRC64;

Query Match 25.8%; Score 54; DB 1; Length 1092;
Best Local Similarity 40.9%; Pred. No. 22;
Matches 9; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 13 YVGAKIGVDKQINGKNTAYG 34
Db 692 FTGAQVKLPCLEISGSVTAYG 713

RESULT 10
MOO_MYCTU STANDARD; PRT; 493 AA.
ID MOO_MYCTU
AC O05807;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PROBABLE MALATE:OUTINONE OXIDOREDUCTASE (EC 1.1.99.16) (MALATE
DE DEHYDROGENASE [ACCEPTOR]) (MOO).
GN RV2852C OR MTC124AL.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekla E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).

```

```

CC -1- CATALYTIC ACTIVITY: (S)-MALATE + ACCEPTOR = OXALOACETATE + REDUCED
CC ACCEPTOR.
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- PATHWAY: CITRIC ACID CYCLE.
CC -1- SIMILARITY: BELONGS TO THE MOO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 295207; CAB08454.1; -.
DR TOBERCULIST; RV2852C; -.
KW Oxidoreductase; FAD.
SQ SEQUENCE 493 AA; 53595 MW; C3B4C469FAF285F CRC64;

Query Match
Best Local Similarity 25.6%; Score 53.5; DB 1; Length 493;
Matches 14; Conservative 8; Mismatches 13; Indels 3; Gaps 2;

OY 4 YGNSADQPYVGAKIGOVDAKOINGKN-TAYGIYAGYN 40
Db 287 YGFPAAPGAPPLGAL--HLDLRFVNGKSMVFGPYAGWS 322

RESULT 11
UREG_HELPEY STANDARD; PRT; 199 AA.
AC 009066;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UREASE ACCESSORY PROTEIN UREG.
GN UREG OR HP0068.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=85P;
RX MEDLINE; 92210488.
RA Cussac V., Ferrero R.L., Labigne A.;
RT *Expression of Helicobacter pylori urease genes in Escherichia coli
RT grown under nitrogen-limiting conditions.*;
RL J. Bacteriol. 174:2466-2473(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA Tomb J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khaliq H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT *The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.*;
RL Nature 388:539-547(1997).
CC -1- FUNCTION: PROBABLY FACILITATING NICKEL INCORPORATION.
CC -1- SIMILARITY: BELONGS TO THE UREG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M84338; AAA25025.1; -.
DR EMBL: A24198; CAA01726.1; -.
DR EMBL: AE000528; AAD07131.1; -.
DR PIR: D41834; D41834.
DR TIGR: HP0068; -.
DR PFAM: PF01495; HypB_Ureg; 1.
KW Nickel; ATP-binding.
FT NP_BIND 8 15 ATP (POTENTIAL).
FT CONFLICT 102 102 D -> S (IN REF. 1).
FT CONFLICT 167 176 MRGEKPIFT -> IAKSPFIIP (IN REF. 1).
SQ SEQUENCE 199 AA; 21955 MW; A23FE448EB6A208B1 CRC64;

Query Match
Best Local Similarity 24.4%; Score 51; DB 1; Length 199;
Matches 11; Conservative 6; Mismatches 7; Indels 2; Gaps 1;

OY 6 NSADADQPYVGA--KIGOVDAKOINGK 29
Db 145 NKIDLAPYVGADLKVMERDSKMRGE 170

RESULT 12
YHIO_NEIGO STANDARD; PRT; 250 AA.
ID YHIO_NEIGO
AC P72077;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 27.3 KD PROTEIN.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11;
RA Carrick C.S., Pyfe J.A.M., Davies J.K.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: STRONG, TO E.COLI YHIO AND H.INFLUENZAE HI0949.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U65994; AAC82508.1; -.
DR Hypothetical protein.
KW SEQUENCE 250 AA; 27316 MW; FC749DB91985763 CRC64;

Query Match
Best Local Similarity 24.2%; Score 50.5; DB 1; Length 250;
Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

OY 11 QPYVGAKI-GOVDAKOINGKNTAYGIYAGY 39
Db 216 RPRUGEHLAGAPAYQYTGKSTRFDVILPY 245

RESULT 13
OPA_HAEIN STANDARD; PRT; 121 AA.
ID OPA_HAEIN
AC P45086;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE OPA PROTEIN.
GN OPA OR H11174.

```

```

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20.
RX MEDLINE: 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Balt C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shrivley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich L.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT *Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE OF 39-121 FROM N.A.
RC STRAIN-RD / KW20:
RX MEDLINE: 96134971.
RA Preston A., Maskell D., Johnson A., Moxon E.R.;
RT *Altered lipopolysaccharide characteristic of the 169 phenotype in
RT Haemophilus influenzae results from mutations in a novel gene, 'lsn.';
RL J. Bacteriol. 178:396-402(1996).
CC -1- SIMILARITY: SOME TO N.GONORRHOEA OPACITY PROTEIN OPA66.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32797; AAC2827.1; -
DR EMBL: U17295; AAA95987.1; -
DR TIGR: H11174; -
DR SEQUENCE 121 AA; 13153 MW; 6647D8B471F42223 CRC64;

Query Match 23.9%; Score 50; DB 1; Length 121;
Best Local Similarity 26.7%; Pred. No. 8.7;
Matches 12; Conservative 6; Mismatches 17; Indels 10; Gaps 1;

QY 4 YGNSADAQPYVGAKIGVDKQINGK-----NTATGCIYAG 38
Db 36 FDNISKVQPYVGARVATNQPKYTNRAEQKPKSSSDIKLGYGVAG 80

RESULT 14
Y05F_BP74 STANDARD; PRT; 325 AA.
ID Y05F_BP74
AC P39261; Q96218;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOPHARYNGEAL 36.7 KD PROTEIN IN NRDC-MOEB INTERGENIC REGION.
GN Y05F OR NRDC.10.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
RN [1]
RP SEQUENCE FROM N.A.
RA Mzhavia N., Marusich E., Djavakhishvili T., Netzel J., Peterson S.,
RA Awaysa M., Eldermiller J., Canada D., Tracy J., Galbreath K.,
RA Paddison P., Anderson B., Stidham T., Blatner F., Kutter E.M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U76612; AAB26972.1; -
KW Hypothetical protein.
SQ SEQUENCE 325 AA; 36687 MW; FD6E3A67A3D403BF CRC64;

Query Match 23.9%; Score 50; DB 1; Length 325;
Best Local Similarity 47.4%; Pred. No. 23;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 18 IGQVDKQINGKMTAYGIY 36
Db 126 IGIEYKISNGKISGIGLY 144

RESULT 15
RUXG_SCHPO STANDARD; PRT; 77 AA.
ID RUXG_SCHPO
AC 074966;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROBABLE SMALL NUCLEAR RIBONUCLEOPROTEIN G (SNRNP-G) (SM PROTEIN G).
GN SPBC84.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Beck A., Reinhardt R., Lyne M., Wood V., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE COMMON SM PROTEIN, IS FOUND IN U1 AND U2
CC SNRPS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL023706; CAA19285.1; -
DR PFAM: PF01423; Sm; 1.
KW Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing.
SQ SEQUENCE 77 AA; 8604 MW; 6468E36CEFA8139A CRC64;

Query Match 23.7%; Score 49.5; DB 1; Length 77;
Best Local Similarity 32.4%; Pred. No. 6.5;
Matches 11; Conservative 4; Mismatches 10; Indels 9; Gaps 1;

QY 16 AKIGVDK-----QINGKMTAYGIYGVN 40
Db 2 SKAGAPDLKRYLDRQVYVQVNGSRKRYGVLRGID 35

```

Search completed: May 20, 2000, 12:22:44
Job time: 3958 sec

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	63.5	30.4	407	2	P96774	p96774 haemophilus
2	59.5	28.5	380	5	O9XRP9	O9XRP9 caenorhabditis
3	56	28.8	346	2	O51641	O51641 actinobacillus
4	55.5	26.6	767	3	O12357	O12357 saccharomyces
5	55.5	26.6	1155	3	O07163	O07163 saccharomyces
6	55.5	26.6	1196	3	O13527	O13527 saccharomyces
7	55.5	26.6	1285	3	O09334	O09334 saccharomyces
8	55.5	26.6	1328	3	O99231	O99231 saccharomyces
9	55.5	26.6	1348	3	O03855	O03855 saccharomyces
10	55.5	26.6	1346	3	O04345	O04345 saccharomyces
11	55.5	26.6	1346	3	O05679	O05679 saccharomyces
12	55.5	26.6	1346	3	O05369	O05369 saccharomyces
13	55.5	26.6	1347	3	O03494	O03494 saccharomyces
14	55.5	26.6	1759	3	O99337	O99337 saccharomyces
15	55.5	26.6	1755	3	O12088	O12088 saccharomyces
16	55.5	26.6	1755	3	O12112	O12112 saccharomyces
17	55.5	26.6	1755	3	O12141	O12141 saccharomyces
18	55.5	26.6	1755	3	O12269	O12269 saccharomyces
19	55.5	26.6	1755	3	O12273	O12273 saccharomyces
20	55.5	26.6	1755	3	O12316	O12316 saccharomyces

RESULT	1	ALIGNMENTS
P96774	P96774	P96774
AC	P96774	P96774
DT	01-MAY-1997 (TREMBLrel. 03, Created)	
DT	01-MAY-1997 (TREMBLrel. 03, Last sequence update)	
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)	
DE	OMP42.	
GN	OMP42.	
OS	Haemophilus ducreyi.	
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;	
OC	Haemophilus.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-35000;	
RX	MEDLINE: 97197543.	
RA	KLEINER-TAIT J., HILDE T.J., MACIVER I., SPINOLA S.M., RADOLF J.D.,	
RA	HANSEN E.J.,	
RT	"The major outer membrane protein of Haemophilus ducreyi consists of	
RT	two OmpA homologs."	
RL	J. Bacteriol. 179:1764-1773(1997).	
DR	EMBL: U60646; AAB49274.1; -.	
DR	HSSP: P02934; 1BXW.	
DR	PFAM: PF00691; OmpA. 1.	
DR	PFAM: PF01389; OmpA_membrane; 1.	
DR	PRINTS: PRO1021; OMPADOMAIN.	
SO	SEQUENCE 407 AA; 44702 MW; 4C2BA5D7 CRC32;	
QY	8 ADAQP-----YVGAKIG-----QVAKINGKNTA-----YGIYAG 38	
Db	17 ATAAQAQDTFYGARAGASFRHGINQFPNKKTKRKHTAVMEGKLNIRDSVTVGVAG 76	
QY	39 Y 39	
Db	77 Y 77	
Query Match	30.4%; Score 63.5; DB 2; Length 407;	
Best Local Similarity	34.4%; Pred. No. 1.2;	
Matches	21; Conservative 3; Mismatches 8; Indels 29; Gaps 3	

```

O9XTP9
ID O9XTP9 PRELIMINARY; PRT; 380 AA.
AC O9XTP9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE ZK1321.4 PROTEIN.
GN ZK1321.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditida;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA GARDNER A.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z48717; CAA88612.1; -.
DR EMBL; Z48584; CAA88612.1; JOINED.
DR EMBL; Z48584; CAA88478.1; -.
DR EMBL; Z48717; CAA88478.1; JOINED.
SQ SEQUENCE 380 AA; 40259 MW; 1D99FAD9 CRC32;

Query Match 28.5%; Score 59.5; DB 5; Length 380;
Best Local Similarity 38.6%; Pred. No. 3.7;
Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

QY 3 SYGNSADAPYVGAIGQVDAKQIN-----GKNTAYGIVAG 38
Db 224 SYNGNATSPY-GAGSGGTPLNMTFTINTSPAGNCATYGAQCG 266

RESULT 3
O51841 PRELIMINARY; PRT; 346 AA.
AC O51841;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE OUTER MEMBRANE PROTEIN 34 PRECURSOR.
GN OMP34.
OS Actinobacillus actinomycetemcomitans
OS (Haemophilus actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NCTC 9710;
RX MEDLINE; 96084499.
RA WHITE P.A., NAIR S.P., KIM M.J., WILSON M., HENDERSON B.;
RT "Molecular characterization of an outer membrane protein of Actinobacillus actinomycetemcomitans belonging to the Ompa family."
RL Infect. Immun. 66:369-372(1998).
DR EMBL; AF005079; AAC00068.1; -.
DR HSP; P02934; IBXW.

```

```

DR PROSITE; PS01068; OMPA; 1.
DR PFAM; PF00691; OMPA; 1.
DR PFAM; PF01389; OMPA_membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
KW Signal.
FT SIGNAL. 1 21 POTENTIAL.
SQ SEQUENCE 346 AA; 36905 MW; 8AB0E0A3 CRC32;

Query Match 26.8%; Score 56; DB 2; Length 346;
Best Local Similarity 35.4%; Pred. No. 9.5;
Matches 17; Conservative 4; Mismatches 11; Indels 16; Gaps 3;

QY 8 ADAQP-----YVGAIGQVDA-----KQ-----INGKNTAYGIVAGY 39
Db 19 AQAAPQANTFYAGAKAGKAGMASSHHGLNQFKQGVSIINNSEAYGVFGY 66

RESULT 4
O12357 PRELIMINARY; PRT; 767 AA.
AC O12357;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE 01-NOV-1996 (TREMBLrel. 01, last annotation update)
DE FRAME-SHIFT IN TYB PROBABLY NOT FUNCTIONAL.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA RIEGER M., MUELLER-AUER S., BRUECKNER M., SCHAEFER M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MIPS.
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z72824; CAA97037.1; -.
DR EMBL; Z72823; CAA97029.1; -.
SQ SEQUENCE 767 AA; 87951 MW; B97D40F4 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 767;
Best Local Similarity 43.8%; Pred. No. 29;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAIGQV---DAKQINGKNT 31
Db 624 SYGN---QPYKSOIGINIFLNGKVIYGKST 651

RESULT 5
O07163 PRELIMINARY; PRT; 1155 AA.
AC O07163;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE (STRAIN JB84A CONTAINING PLASMID PNN162).
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-JB84A;
RX MEDLINE; 88246410.
RA BOEKE J.D., EICHINGER D., CASTRILLON D., FINK G.R.;
RT "The Saccharomyces cerevisiae genome contains functional and nonfunctional copies of transposon Ty1."
RL Mol. Cell. Biol. 8:1432-1442(1988).
DR EMBL; M18706; AAA66938.1; -.

```


SQ SEQUENCE 1155 AA; 131935 MW; 41599191 CRC32;
 Query Match 26.6%; Score 55.5; DB 3; Length 1155;
 Best Local Similarity 43.8%; Pred. No. 47;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
 Oy 3 SYGNSADAPYVAGKIGOV---DAKQINGKNT 31
 Db 1012 SYGN---QPYKSOIGNITLNGKVIQKST 1039
 ||||| ||| ::||| : |||||
 RESULT 6
 ID 013527 PRELIMINARY; PRT; 1196 AA.
 AC 013527;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JUN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE VAR009CP.
 GN VAR009C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 RN Saccharomycetaceae; Saccharomyces.
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 85182712.
 RA DRABKIN H.J., RAJBHANDARY U.L.;
 RT "Attempted expression of a human initiator tRNA gene in Saccharomyces
 cerevisiae.";
 RL J. Biol. Chem. 260:5596-5602(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 86028187.
 RA HIETER P., PRIDMORE D., HEGEMANN J.H., THOMAS M., DAVIS R.W.,
 RA PHILIPPSEN P.;
 RT "Functional selection and analysis of yeast centromeric DNA.";
 RL Cell 42:913-921(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 89128457.
 RA SLATER M.R., CRAIG E.A.;
 RT "The SSA1 and SSA2 genes of the yeast Saccharomyces cerevisiae.";
 RL Nucleic Acids Res. 17:805-806(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 90291999.
 RA HEYER W.D., RAO M.R., ERDILE L.F., KELLY T.J., KOLODNER R.D.;
 RT "An essential Saccharomyces cerevisiae single-stranded DNA binding
 protein is homologous to the large subunit of human RP-A.";
 RL EMBO J. 9:2321-2329(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 91071607.
 RA WYTE W., KROPP L.H., LAMB J., CROWLEY J.C., KABACK D.B.;
 RT "Molecular cloning of chromosome I DNA from Saccharomyces cerevisiae:
 isolation, characterization and regulation of the SP07 sporulation
 gene.";
 RL Gene 95:65-72(1990).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 91353080.
 RA SCHWETZER B., PHILIPPSEN P.;
 RT "CDC15, an essential cell cycle gene in Saccharomyces cerevisiae,
 encodes a protein kinase domain.";
 RL Yeast 7:265-273(1991).
 RN [7]

RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 92051323.
 RA DAVIES C.J., HUTCHISON C.A. III.;
 RT "A directed DNA sequencing strategy based upon Tr3 transposon
 mutagenesis: application to the ADE1 locus on Saccharomyces cerevisiae
 chromosome I.";
 RL Nucleic Acids Res. 19:5731-5738(1991).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 93066269.
 RA LEEFEBRE O., CARLES C., CONESA C., SWANSON R.N., BOUET F., RIVA M.,
 RA SENTENAC A.;
 RT "TFC3: gene encoding the B-block binding subunit of the yeast
 transcription factor TFC2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10512-10516(1992).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 9311037.
 RA HIRAGA K., SUZUKI K., TSUCHIYA E., MIYAKAWA T.;
 RT "Cloning and characterization of the elongation factor EF-1 beta
 homologue of Saccharomyces cerevisiae. EF-1 beta is essential for
 growth.";
 RL FEBS Lett. 316:165-169(1993).
 RN [10]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 95028152.
 RA CLARK M.W., KENG T., STORMS R.K., ZHONG W., FORTIN N., ZENG B.,
 RA DELANEY S., OUELLETTE B.F., BARTON A.B., AND KABACK D.B.;
 RT "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
 the 42 kbp SP07-CEN1-CDC15 region.";
 RL Yeast 10:535-541(1994).
 RN [11]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 95249563.
 RA BUSSEY H., KABACK D.B., ZHONG W., VO D.T., CLARK M.W., FORTIN N.,
 RA HALL J., OUELLETTE B.F., KENG T., BARTON A.B., SU Y., DAVIES C.K.,
 RA STORMS R.K.;
 RT "The nucleotide sequence of chromosome I from Saccharomyces
 cerevisiae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX JIA Y., CHERRY J.M.;
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; L22015; AAC04967.1; -
 SQ SEQUENCE 1196 AA; 136731 MW; BA97ERCF CRC32;
 Query Match 26.6%; Score 55.5; DB 3; Length 1196;
 Best Local Similarity 43.8%; Pred. No. 49;
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
 Oy 3 SYGNSADAPYVAGKIGOV---DAKQINGKNT 31
 Db 1053 SYGN---QPYKSOIGNITLNGKVIQKST 1080
 ||||| ||| ::||| : |||||
 RESULT 7
 ID 003934 PRELIMINARY; PRT; 1285 AA.
 AC 003934;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE HYPOTHETICAL PROTEIN (FRAGMENT).
 GN TYB.
 OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA MURPHY L., HARRIS D.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z68329; CAA92721.1; -
KW PROSITE; PS00141; ASP_PROTEASE; 1.
FT Hypothetical protein; Hydrolase; Aspartyl protease.
FT NON_TER 1285 1285
SQ SEQUENCE 1285 AA; 146556 MW; 03F439B7 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1285;
Best Local Similarity 43.8%; Pred. No. 54;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31
Db 1203 SYGN---QPYKSOIGNIFLNGKVIYGKST 1230

RESULT 8
O99231 PRELIMINARY; PRT; 1328 AA.
AC O99231;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE HYPOTHEICAL 151.2 KD PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA OLIVER K., SHORE L., HARRIS D.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z68194; CAA92352.1; -
DR EMBL; Z68195; CAA92360.1; -
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Aspartyl protease.
SQ SEQUENCE 1328 AA; 151193 MW; 094C7A55 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1328;
Best Local Similarity 43.8%; Pred. No. 56;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSOIGNIFLNGKVIYGKST 1212

RESULT 9
003855

ID 003855 PRELIMINARY; PRT; 1328 AA.
AC 003855;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE TYB PROTEIN.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA MURPHY L., HARRIS D.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BARRELL B., RAJANDREAM M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z47746; CAA87673.1; -
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Hydrolase; Aspartyl protease.
SQ SEQUENCE 1328 AA; 151007 MW; AC60FC55 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1328;
Best Local Similarity 43.8%; Pred. No. 56;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSOIGNIFLNGKVIYGKST 1212

RESULT 10
O04345 PRELIMINARY; PRT; 1346 AA.
AC O04345;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE HYPOTHEICAL 153.9 KD PROTEIN.
GN TYB, YD9673.05C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA CONNOR R., CHURCHER C.M.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z68196; CAA92372.1; -
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153889 MW; E0D66880 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1346;
Best Local Similarity 43.8%; Pred. No. 57;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31
Db 1203 SYGN---QPYKSOIGNIFLNGKVIYGKST 1230

RESULT 11
005679

ID 005679 PRELIMINARY; PRT: 1346 AA.
AC 005679;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE TY B.
GN YBL0822.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RX MEDLINE: 96076635.
RA OBERMAIER B., GASSENHUBER J., PIRAVANDI E., DOMDEY H.;
RT "Sequence analysis of a 78.6 kb segment of the left end of
RT Saccharomyces cerevisiae chromosome II.";
RL Yeast 11:1103-1112(1995).
DR EMBL: X79489; CAA55998.1; -;
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153931 MW; 68F2AB6B CRC32;
Query Match 26.6%; Score 55.5; DB 3; Length 1346;
Best Local Similarity 43.8%; Pred. No. 57;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
QY 3 SYGNSADAOPYGAKIGOV---DAKOINKNT 31
Db 1203 SYGN---QPYKSOIGNIFLNGKVGKST 1230
RESULT 12
005369 PRELIMINARY; PRT: 1346 AA.
ID 005369;
AC 005369;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE TRANSOSON TY1-17 154.0KD HYPOTHETICAL PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RX MEDLINE: 95400292.
RA MURAKAMI Y., NAITOU M., HAGIMARA H., SHIBATA T., OZAWA M.,
RA SASANUMA S., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,
RA YAMAZAKI M., TASHIRO H., EKI T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
DR EMBL: D50617; BAA09237.1; -;
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153890 MW; 1A8EBF4B CRC32;
Query Match 26.6%; Score 55.5; DB 3; Length 1346;
Best Local Similarity 43.8%; Pred. No. 57;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
QY 3 SYGNSADAOPYGAKIGOV---DAKOINKNT 31
Db 1203 SYGN---QPYKSOIGNIFLNGKVGKST 1230
RESULT 13
003494 PRELIMINARY; PRT: 1347 AA.
ID 003494

AC 003494;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 153.9 KD PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA OLIVER K., SHORE L., HARRIS D.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z68194; CAA92351.1; -;
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hypothetical protein; Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1347 AA; 153944 MW; DEA98B36 CRC32;
Query Match 26.6%; Score 55.5; DB 3; Length 1347;
Best Local Similarity 43.8%; Pred. No. 57;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
QY 3 SYGNSADAOPYGAKIGOV---DAKOINKNT 31
Db 1204 SYGN---QPYKSOIGNIFLNGKVGKST 1231
RESULT 14
009337 PRELIMINARY; PRT: 1749 AA.
ID 009337;
AC 009337;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CHROMOSOME XIV READING FRAME ORF YNL055C.
GN TY1B.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA BERGEE P., DOIGNON F., CROUZET M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z71331; CAA95928.1; -;
DR EMBL: Z71330; CAA95924.1; -;
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PFAM: PF01021; TYA; 1.
KW Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1749 AA; 198196 MW; 140C9F2E CRC32;
Query Match 26.6%; Score 55.5; DB 3; Length 1749;
Best Local Similarity 43.8%; Pred. No. 78;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
QY 3 SYGNSADAOPYGAKIGOV---DAKOINKNT 31
Db 1606 SYGN---QPYKSOIGNIFLNGKVGKST 1633
RESULT 15
012088

```

ID 012088      PRELIMINARY;      PRT; 1755 AA.
AC 012088;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE CHROMOSOME XII READING FRAME ORF YLR036C.
GN TY1B.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP PARTIAL SEQUENCE FROM N.A.
RA OBERMAIER B., PIRAVANDI E., RINKE M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-0 FROM N.A.
RA KOETTER P., ROSE M., ENTIAN K.D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 273208; CAA97563.1; -.
DR EMBL; 273207; CAA97560.1; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFAM; PF01021; TYA; 1.
KW Hydrolase; Aspartyl protease.
SQ SEQUENCE 1755 AA; 198460 MW; 41619FC4 CRC32;

Query Match      26.68; Score 55.5; DB 3; Length 1755;
Best Local Similarity 43.8%; Pred. No. 78;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADADPYVGAKIGV--DAKQINGKNT 31
Db 1612 SYGN---QPYRKSQIGNITYLNGKVIQKST 1639

```

Search completed: May 20, 2000, 12:22:02
 Job time: 4019 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 11:01:59 ; Search time 1501.5 Seconds
(without alignments)
-351.799 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543
Sequence: 1 atgaacttaataaacact.....gcgcctattgcttttaa 543

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl1:
1: gb_da1:*
2: gb_da2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pl4:*
11: gb_pl5:*
12: gb_pl6:*
13: gb_pl7:*
14: gb_pl8:*
15: gb_pl9:*
16: gb_pl10:*
17: gb_pl11:*
18: gb_pl12:*
19: gb_pl13:*
20: gb_pl14:*
21: gb_pl15:*
22: gb_pl16:*
23: gb_pl17:*
24: gb_pl18:*
25: gb_pl19:*
26: gb_pl20:*
27: gb_pl21:*
28: gb_pl22:*
29: gb_pl23:*
30: gb_pl24:*
31: gb_pl25:*
32: gb_pl26:*
33: gb_pl27:*
34: gb_pl28:*
35: gb_pl29:*
36: gb_pl30:*
37: gb_pl31:*
38: gb_pl32:*
39: gb_pl33:*
40: gb_pl34:*
41: gb_pl35:*
42: gb_pl36:*
43: gb_pl37:*
44: gb_pl38:*
45: gb_pl39:*

45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pl4:*
52: gb_pl5:*
53: gb_pl6:*
54: gb_pl7:*
55: gb_pl8:*
56: gb_pl9:*
57: gb_pl10:*
58: gb_pl11:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	8.3	1164	34	TTTAP1A
2	45	8.3	1575	34	TTTAP1N
3	44.4	8.2	119118	11	HSRRY7CCL
4	43.6	8.0	823	7	CAMODUT1
5	43.6	8.0	122852	41	AC009770
6	43.6	8.0	189286	40	AC008125
7	43.6	8.0	208437	57	AC018655
8	42.2	7.8	308399	44	AC005140
9	41.6	7.7	1302	51	AF090946
10	41.4	7.6	172633	45	AC016048
11	40.8	7.5	1277	7	SCU20616
12	40.2	7.4	10976	2	AE001127
13	39.4	7.3	846	5	128501
14	39.4	7.3	873	2	AF077664
15	39.4	7.3	891	5	128499
16	39.4	7.3	934	1	BBO5PBB31
17	39.4	7.3	934	1	BBO5PBB
18	39.4	7.3	934	1	BBO5PBB
19	39.4	7.3	1653	1	BOROSPAC
20	39.4	7.3	1653	1	BOROSPAC
21	39.4	7.3	1653	1	BOROSPAC
22	39.4	7.3	1653	1	BOROSPAC
23	39.4	7.3	1653	1	BOROSPAC
24	39.4	7.3	1653	1	BOROSPAC
25	39.4	7.3	1653	1	BOROSPAC
26	39.4	7.3	1915	1	BROSPAB
27	39.4	7.3	1915	5	A04009
28	39.4	7.3	1916	1	BOROSPAB
29	39.4	7.3	1916	1	BOROSPAB
30	39.4	7.3	1916	1	BOROSPAB
31	39.4	7.3	1916	1	BOROSPAB
32	39.4	7.3	1916	1	BOROSPAB
33	39.4	7.3	1916	1	BOROSPAB
34	39.4	7.3	1916	1	BOROSPAB
35	39.4	7.3	1916	1	BOROSPAB
36	39.4	7.3	1916	1	BOROSPAB
37	39.4	7.3	1916	1	BOROSPAB
38	39.4	7.3	1916	1	BOROSPAB
39	39.4	7.3	1916	1	BOROSPAB
40	39.4	7.3	1916	1	BOROSPAB
41	39.4	7.3	1916	1	BOROSPAB
42	39.4	7.3	1916	1	BOROSPAB
43	39.4	7.3	1916	1	BOROSPAB
44	39.4	7.3	1916	1	BOROSPAB
45	39.4	7.3	1916	1	BOROSPAB

ALIGNMENTS

[illegible][illegible]

Query Match	8.3%	Score 45	DB 34	Length 1575
Best Local Similarity	50.7%	Pred. No. 0.2		
Matches 108	Conservative 0	Mismatches 105	Indels 0	Gaps 0
OY	83	attctcgtatgctcaaccctatgttggctgcaaaattgtcgaagtacgaccgaagcaaa	142	
DB	1064	ATGGATGATGATGCTGCAACGCGTGATGATGCCGAAGATGATGATGATTAACGAAGATGCCG	1123	
OY	143	tcaacggtgaagaacacgcgcctatggtatcttatgcgaagttacacttgaccaaaatttg	202	
DB	1124	AAGATGGTATGATGATGCTCAAGATGCTGATGATGCCGAAGATGATGATGATTAACGAAGATG	1183	
OY	203	gcgtagaacccgaatttgcttgcacgcgcgaagaatttaatgcgaagcgttgacctg	262	
DB	1184	GGATGATGATGCCGAAGATGATGATGATGCTGGAAGACGGTATGATGATCCGAAGATGATG	1243	
OY	263	taaaagtgatgtgaagtcctttgtgtgctatg	295	
DB	1244	ATAAGCAAGATGCCGAAGATGATGATGATGCTG	1276	
RESULT 3				
LOCUS	HSRY7CCL1	119118 bp	DNA	PRI 12-DEC-1999
DEFINITION	Human DNA sequence from clone XX-PRY7CCL1 on chromosome 22 Contains an STS, GSSS, genomic marker D22S928, tc and ca repeat polymorphisms and a putative Cpg Island, complete sequence.			
ACCESSION	AL049760			
VERSION	AL049760.26	GI:5777587		
KEYWORDS	HTG; ca repeat polymorphism; Cpg island; D22S928; tc repeat polymorphism.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 119118)			
TITLE	Ramsay H.			
JOURNAL	Direct Submission			
COMMENT	Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Requests: clonerequests@sanger.ac.uk On Aug 26, 1999 this sequence version replaced gi:5763790. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information http://www.sanger.ac.uk/Projects/C_elegans/wormpep IMPORTANT: This sequence is not the entire insert of clone XX-PRY7CCL1 it may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone XX-PRY7CCL1 is at 1 in this sequence. The true left end of clone CTA-217C2 is at 119019 in this sequence. The true right end of clone RP4-753M9 is at 50597 in this sequence.			

FEATURES	Location/Qualifiers
source	1. .119118 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /clone="XX-PRYR/CC1" 12. .3186 /note="CpG island" evidence=not_experimental
misc_feature	77. .504 /note="214 copies 2 mer gg 55 conserved"
repeat_region	111. .500 /note="15 copies 26 mer 56 conserved"
repeat_region	369. .488 /note="3 copies 40 mer 76 conserved"
repeat_region	1585. 1620 /note="6 copies 6 mer gtgagt 86 conserved"
repeat_region	2485. .2644 /note="4 copies 40 mer 70 conserved"
repeat_region	2527. .2598 /note="12 copies 6 mer cccccc 69 conserved"
misc_feature	2593. 2608 /note="Single clone region"
misc_feature	2647. .2669 /note="Single clone region"
misc_feature	2719. .2726 /note="Single clone region"
repeat_region	2744. 2817 /note="37 copies 2 mer cc 66 conserved"
repeat_region	3678. .3707 /note="15 copies 2 mer tt 90 conserved"
repeat_region	3956. .4249 /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region	4418. 4542 /note="AluDo/FRAM repeat: matches 152. .276 of consensus"
repeat_region	5264. .5329 /note="MIR repeat: matches 95. .165 of consensus"
repeat_region	5399. .6097 /note="MIR repeat: matches 70. .172 of consensus"
repeat_region	6121. 6177 /note="MER94 repeat: matches 47. .107 of consensus"
repeat_region	6178. .6472 /note="AluSg repeat: matches 1. .295 of consensus"
repeat_region	6473. .6500 /note="MER94 repeat: matches 27. .47 of consensus"
repeat_region	6501. 6797 /note="AluY repeat: matches 1. .297 of consensus"
repeat_region	7334. .7400 /note="MER21B repeat: matches 706. .789 of consensus"
repeat_region	7179. .7894 /note="MER2 repeat: matches 18. .132 of consensus"
repeat_region	7895. 7975 /note="AluY repeat: matches 2. .82 of consensus"
repeat_region	8005. .8080 /note="MER2 repeat: matches 269. .343 of consensus"
repeat_region	8105. .8229 /note="AluSg/x repeat: matches 9. .133 of consensus"
repeat_region	8257. .8415 /note="AluDo/FRAM repeat: matches 149. .304 of consensus"
repeat_region	8630. .8919 /note="AluJo repeat: matches 13. .300 of consensus"
repeat_region	9266. .9580 /note="AluSp repeat: matches 5. .313 of consensus"
repeat_region	9714. .9845 /note="MIR repeat: matches 105. .262 of consensus"
repeat_region	9973. .10170 /note="MIR repeat: matches 34. .240 of consensus"
repeat_region	10270. 10480 /note="L1MA10 repeat: matches 6135. .6322 of consensus"
repeat_region	10481. .10735 /note="AluSg repeat: matches 57. .311 of consensus"
repeat_region	10736. .10821 /note="L1MA10 repeat: matches 6045. .6135 of consensus"

```
repeat_region 10871..11171
/note="Aluuo repeat: matches 1. .303 of consensus"
repeat_region 11190..11487
/note="Alusx repeat: matches 1. .298 of consensus"
repeat_region 11610..11726
/note="MIR repeat: matches 13. .134 of consensus"
repeat_region 12322..12636
/note="Alusx repeat: matches 1. .312 of consensus"
repeat_region 13016..13133
/note="L2 repeat: matches 2616. .2748 of consensus"
repeat_region 13419..13704
/note="143 copies 2 mer gg 59 conserved"
repeat_region 13432..13769
/note="13 copies 26 mer 57 conserved"
repeat_region 13878..14208
/note="MLTIC repeat: matches 6. .368 of consensus"
repeat_region 14224..14886
/note="L1PA10 repeat: matches 5476. .6165 of consensus"
repeat_region 14893..15204
/note="Alusp repeat: matches 1. .312 of consensus"
repeat_region 15226..15335
/note="L1P3 repeat: matches 5336. .5465 of consensus"
repeat_region 15336..15446
/note="MLTIC repeat: matches 355. .464 of consensus"
repeat_region 15484..15777
/note="Alusq repeat: matches 1. .295 of consensus"
misc_feature complement(15768..16376)
/note="match: GSS: Em:AQ476535"
repeat_region 16387..16541
/note="MIR repeat: matches 73. .226 of consensus"
repeat_region 17071..17122
/note="26 copies 2 mer aa 73 conserved"
repeat_region 17222..17308
/note="L2 repeat: matches 2652. .2738 of consensus"
repeat_region 17452..17487
/note="18 copies 2 mer tg 91 conserved"
repeat_region 17599..17908
/note="Alur repeat: matches 1. .310 of consensus"
repeat_region 18508..18616
/note="L1PA6 repeat: matches 6035. .6143 of consensus"
repeat_region 18668..18965
/note="Alusx repeat: matches 1. .298 of consensus"
repeat_region 19069..19100
/note="MLT1-INTERNAL repeat: matches 692. .723 of consensus"
repeat_region 19494..19849
/note="MLT1-INTERNAL repeat: matches 927. .1296 of consensus"
repeat_region 19869..20149
/note="Aluuo repeat: matches 18. .288 of consensus"
repeat_region 20218..20385
/note="MLT1B repeat: matches 1. .180 of consensus"
repeat_region 20386..20684
/note="Alusp repeat: matches 1. .303 of consensus"
repeat_region 20685..20736
/note="MLT1B repeat: matches 180. .234 of consensus"
repeat_region 20737..21016
/note="Alusg repeat: matches 1. .280 of consensus"
repeat_region 21017..21215
/note="MLT1B repeat: matches 234. .390 of consensus"
repeat_region 21217..21347
/note="MIR repeat: matches 12. .144 of consensus"
repeat_region 21582..21731
/note="L2 repeat: matches 2511. .2657 of consensus"
repeat_region 22386..22415
/note="15 copies 2 mer ac 93 conserved"
repeat_region 23579..23881
/note="Alusx repeat: matches 1. .310 of consensus"
misc_feature complement(23985..24510)
/note="match: GSS: Em:AQ529159"
misc_feature complement(24086..24604)
/note="match: GSS: Em:AQ569063"
repeat_region 24115..24201
```

```
repeat_region 24201..24304
/note="MLT1D repeat: matches 6. .94 of consensus"
misc_feature complement(24215..24582)
/note="match: GSS: Em:AQ375812"
misc_feature 24487..24919
/note="match: GSS: Em:AQ206604"

Query Match      8.2%: Score 44.4; DB 11; Length 119118;
Best Local Similarity 49.2%: Pred. No. 0.29;
Matches 117; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 74 gctatgccaattctgctgctcaacctatgttgytgccaanaatggtcagtagcg 133
      ||| || || || || || || || || || || || || || || || || || ||
DB 73202 GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 73143

QY 134 ccaaggaatacaggtlaagaaacacgcgtatagtattatgcaggttaacttgacc 193
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 73142 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 73083

QY 194 aaaatttggcgtagaaccgaatttgtgttcacagcccaagaattaatgcagcg 253
      || || || || || || || || || || || || || || || || || || ||
DB 73082 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 73023

QY 254 tgaagtcgtaaaaggtgtagtgaagcttcttgggtcgtatggtcacatatcgctataa 311
      || || || || || || || || || || || || || || || || || || ||
DB 73022 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72965

RESULT 4
LOCUS CAMODUT1 823 bp DNA PLN 03-AUG-1995
DEFINITION C.albicans (WO-1) DUT1 gene.
ACCESSION X77925
VERSION X77925.1 GI:457720
KEYWORDS DUT1 gene; dutPase.
SOURCE Candida albicans.
ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Candidaceae; Candida.
REFERENCE 1 (bases 1 to 823)
AUTHORS McIntosh,E.M., Looser,J., Haynes,R.H. and Pearlman,R.E.
TITLE MluI site-dependent transcriptional regulation of the Candida
TITLE albicans dutPase gene
JOURNAL Curr. Genet. 26 (5-6), 415-421 (1994)
MEDLINE 95179795
REMARK Erratum: [[published erratum appears in Curr Genet 1995
REFERENCE Apr;27(5):491]]
AUTHORS 2 (bases 1 to 823)
TITLE Pearlman,R.E.
TITLE Direct Submision
JOURNAL Submitted (28-FEB-1994) R.E. Pearlman, York University, Dept of
Biology, Farquharson Bldg., 4700 Keele St., Downsview, Ontario M3J
1P3, CANADA

FEATURES
source 1..823
location/Qualifiers
organism="Candida albicans"
strain="WO-1"
db_xref="taxon:5476"
chromosome="1"
clone_lib="plasmid YEP352"
clone="pBKS+"
map="sfi I fragment S"
gene 149..746
/gene="DUT1"
misc_feature 149..156
/gene="DUT1"
misc_feature 149..156
/note="potential SCB element"
/gene="DUT1"
misc_feature 197..200
/note="Mlu I site (potential MCB element)"
/gene="DUT1"
TATA_signal
```


Query Match	Best Local Similarity	8.0%: Score 43.6; DB 7; Length 823;	Matches 145; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 67	gccatcagctatggaatctctgctatgctcacaacctatgcttggtgcacaatctgtcaa	126	
Db 510	gccgtgaaacagatgattttcaactgctgctgctgctatgacgtcgtatgctgctgata	569	
QY 127	gtagaagcccaagcaaatcaacggtagaagaccgcctatgattattatgcaggtataac	186	
Db 570	gtgaagacagatgcttttttttaaacatcagtgtaaaagatttgcgaatttaaaagagtgataga	629	
QY 187	tttgccacaatcttggtggttagaccgcgaatttggttggtgaagcccaagaatttaac	246	
Db 630	atttgctcatttggtgattttgcaacaaatcgtcgaatgctgattatcgaagaaatttttcacttgaa	689	
QY 247	gcagagcgtgagtcctcgtataaagtgatgtaagtccttttggtgctatgacacatacgc	306	
Db 690	gaattatgattatgacgacgtgtaagctgtaagctgcttggatctacgctgaagaaactragac	749	
QY 307	tataactcatcaataatccccatttatgcgaagggcaaatgagcatgcttaagactaa	366	
Db 750	tagtacttaattttaaataataataatgaaatttataatattgctgattatataataatcgta	809	
QY 367	gtagatggtaccag 380		
Db 810	tttctctatpaaacag 823		
RESULT 5			
AC009770	122852 bp	DNA	HTG 10-SEP-1999
LOCUS	AC009770	122852 bp	DNA
DEFINITION	Human sapiens, *** SEQUENCING IN PROGRESS ***	83 unordered pieces.	
ACCESSION	AC009770.3	GI:5851699	
VERSION	HTG: HTGS_PHSSEL.		
KEYWORDS	human.		
SOURCE	Human sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
REFERENCE	Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 122852)		
	Munzy D.M., Adams C., Aronson A., Bailey M., Barbara J.,		
	Blankenburg K., Bodota B., Bouck J., Bowie S., Brooks A., Buhay C.,		
	Bunac E., Burkett C., Burrows J., Carter M., Chacko V., Chen Z.,		
	Cox C., David R., Delgado O., Deshazo D., Ding Y., Domah-Rashid N.,		
	Dugan-Rocha S., Durbin K.J., Fernandez C., Ferraguto D.,		
	Forcuna-Tansey J., Frantz P., Ganesh R., Gorrell J.H., Gorrell L.L.,		
	Guerra W., Harris K., Hernandez J., Hodgson A., Hughes M.,		
	Holloway C., Hosak H., Jackson L.E., Jackson L., Jia Y., Jones M.,		
	Kelly S., Kondejewski N., Kong Y., Kovar C., Leal B., Li Z.,		
	Lichtarge O., Liu J., Liu W., Logan O., Lu J., Lucier R.,		
	Martin R., Martinez C., McLeod M.P., Mei G., Morgan M., Morris S.,		
	Nash S., Nelson A., Nguyen R., Nguyen N., Nguyen S., Oswal G.,		
	Pattish B., Paxton S., Payton B., Perez L., Pu L.L., Quiles M.,		
	Reiter D., Rives M., Samuel S., Say J., Scherer S., Shah E.,		
	Shen H., Simon M., Sparks A., Stamps A., Sugarcang R., Tabor P.,		
	Taylor T., Vasquez L., Vinson R., Vo O., Wabeh M., Watlington S.,		
	Weinstock G., Weinstock I.R., Williamson A., Worley K., Wren J.,		

TITLE	Wrensford,G., Yu,W., Zhou,X. and Gibbs,R.
JOURNAL	Direct Submission
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 122852)
TITLE	Morley,R.C.
JOURNAL	Direct Submission
COMMENT	Submitted (01-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Sep 9, 1999 this sequence version replaced g1:5822592. * NOTE: This is a 'working draft' sequence. It currently consists of 83 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
	1
*	3284: contig of 3284 bp in length
*	3285
*	3309: gap of unknown length
*	3310
*	6343: contig of 3034 bp in length
*	6344
*	6368: gap of unknown length
*	6369
*	9396: contig of 3028 bp in length
*	9397
*	9421: gap of unknown length
*	9422
*	12446: contig of 3025 bp in length
*	12447
*	15221: gap of unknown length
*	15222
*	15226: contig of 2750 bp in length
*	15247
*	17899: gap of unknown length
*	17900
*	17924: gap of unknown length
*	17925
*	20312: contig of 2388 bp in length
*	20313
*	20336: gap of unknown length
*	20337
*	22705: contig of 2369 bp in length
*	22706
*	22729: gap of unknown length
*	22730
*	25095: contig of 2366 bp in length
*	25096
*	25119: gap of unknown length
*	25120
*	27456: contig of 2337 bp in length
*	27457
*	27480: gap of unknown length
*	27481
*	29761: contig of 2281 bp in length
*	29762
*	29785: gap of unknown length
*	29786
*	31967: contig of 2162 bp in length
*	31968
*	31991: gap of unknown length
*	31992
*	34129: contig of 2138 bp in length
*	34130
*	34153: gap of unknown length
*	34154
*	36287: contig of 2134 bp in length
*	36288
*	36311: gap of unknown length
*	36312
*	38444: contig of 2133 bp in length
*	38445
*	38468: gap of unknown length
*	38469
*	40566: contig of 2098 bp in length
*	40567
*	40590: gap of unknown length
*	40591
*	42605: contig of 2015 bp in length
*	42606
*	42629: gap of unknown length
*	42630
*	44585: contig of 1956 bp in length
*	44586
*	44609: gap of unknown length
*	44610
*	46497: contig of 1888 bp in length
*	46498
*	46521: gap of unknown length
*	46522
*	48376: contig of 1855 bp in length
*	48377
*	48400: gap of unknown length
*	48401
*	50167: contig of 1767 bp in length
*	50168
*	50191: gap of unknown length
*	50192
*	51933: contig of 1742 bp in length
*	51934
*	51957: gap of unknown length
*	51958
*	53692: contig of 1735 bp in length
*	53693
*	53716: gap of unknown length
*	53717
*	53368: contig of 1652 bp in length
*	53369
*	55392: gap of unknown length
*	55393
*	57008: contig of 1616 bp in length
*	57009
*	57032: gap of unknown length
*	57033
*	58616: contig of 1584 bp in length
*	58617
*	58640: gap of unknown length
*	58641
*	60193: contig of 1533 bp in length
*	60194
*	60217: gap of unknown length
*	60218
*	61768: contig of 1551 bp in length

Holloway, C., Hosak, H., Issar, A., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kneitz, S., Kondrjewski, N., Kong, Y., Kovari, C., Lau, S., Leal, B., Lee, E., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R., Maronde, I., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Merscher, S., Miller, A., Montgomery, K.T., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswald, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Shim, C., Simon, M., Sparks, A., Stamps, A., Sugang, R., Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wahbh, M., Watlington, S., Weinstein, G., Weinstein, K.T., Williamson, A., Worley, K., Wren, J., Wrenford, G., Xiang, A.M., Yang, R., Yu, W., Zhou, X., Kuchelapatti, R., Nelson, D. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 189286)
Morley, K.C.

Direct Submission
Submitted (24-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 189286)
Morley, K.C.

Direct Submission
Submitted (25-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 25, 1999 this sequence version replaced g1:5851701.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 189286
Phrap values in estimate: 188412
Average error rate (BCM-Phrap estimate): 0.000392177
Fraction of Phrap values less than 40 : 0.0664395

Number of consensus changing edits:
Number of N's in consensus : 22
0

Position	Original+Context	Edited+Context
5887	atcacacct(n)caagttccct	atcacacct(c)caagttccct
42337	aatgcttct(n)caagctcagt	aatgcttct(c)caagctcagt
47267	ttttttaga(n)gagctctgc	ttttttaga(t)gagctctgc
55095	tgatgtaagc(n)gtttccatc	tgatgtaagc(c)gtttccatc
74098	caagcttct(n)nnntccctc	caagcttct(g)nnntccctc
74099	agcttctctn(n)ntccctctt	agcttcttg(c)ntccctctt
74100	gcttctctnn(n)ntccctctt	gcttcttg(c)ntccctctt
74101	cttctctnn(n)tcctctttac	cttcttg(c)tcctctttac
77151	tgaagatgg(n)aaagatatc	tgaagatgg(a)aaagatatc
77153	aagaatggn(n)agatatcaa	aagaatgga(a)agatatcaa
95463	ttgaagagcc(n)ccacagttc	ttgaagagcc(a)ccacagttc
103775	gttaacatagg(a)atacagttc	gttaacatgg(t)atacagttc
104001	atgaatggt(n)tagctttac	atgaatggt(c)tagctttac
151027	ccttagaacc(n)cacnaattac	ccttagaacc(c)cacnaattac
151031	agaaacacac(n)aattactaac	agaaacacac(a)aattactaac
151057	ctcaagagga(n)aacgaataac	ctcaagagga(a)aacgaataac
151154	ccaatggtt(n)aagaagaat	ccaatggtt(a)aagaagaat
151643	cccaacagtaa(n)aatcacagag	cccaacagta(a)aatcacagag
173642	tatatatagt(n)tatatatata	tatatatagt(a)tatatatata
173684	atatatctt(n)ctatatata	atatatctt(a)ctatatata
179756	atgatctt(n)tgagacacac	atgatctt(c)tgagacacac
186055	ggagccccc(n)ccaaacactg	ggagccccc(a)ccaaacactg

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
10001					*	*	*	*	*
9001				*	*	*	*	*	*
8001				*	*	*	*	*	*
7001				*	*	*	*	*	*
6001				*	*	*	*	*	*
5001				*	*	*	*	*	*
4001			*	*	*	*	*	*	*
3001			*	*	*	*	*	*	*
2001			*	*	*	*	*	*	*
1001		*	*	*	*	*	*	*	*
01		*	*	*	*	*	*	*	*

Version: 1.01 gxf0.

FEATURES

source

Location/Qualifiers
1..189286
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RPC111-25E2"
58..174
/rpt_family="MER103"
complement(309..376)
/rpt_family="Alu"
1586..1653
/rpt_family="MIR"
complement(1693..1817)
/rpt_family="MIR"
1882..2059
/rpt_family="L2"
complement(2681..2813)
/rpt_family="L2"
complement(3681..3898)
/rpt_family="MIR"
4057..4368
/rpt_family="AluSp"
6005..6025

repeat_region


```
*
* 6660 6679: gap of unknown length
* 6680 7547: contig of 868 bp in length
* 7548 7567: gap of unknown length
* 7568 8375: contig of 808 bp in length
* 8376 8395: gap of unknown length
* 8396 9213: contig of 818 bp in length
* 9214 9233: gap of unknown length
* 9234 9400: contig of 167 bp in length
* 9401 9420: gap of unknown length
* 9421 50801: contig of 41381 bp in length
* 50802 50821: gap of unknown length
* 50822 95943: contig of 45122 bp in length
* 95944 95963: gap of unknown length
* 95964 208437: contig of 112474 bp in length.
*
* Location/Qualifiers
*   1. .208437
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="12q"
*     /clone="RP11-946G22"
*
BASE COUNT 61426 a 43297 c 43806 g 59518 t 390 others
ORIGIN
Query Match 8.0%; Score 43.6; DB 57; Length 208437;
Best Local Similarity 47.7%; Pred. No. 0.46;
Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
QY 56 gtcgaacgctgcacatgacgtatgctgcaattcgtgctgcaacccattgtgtgcca 115
Db 80546 GTGATGATGCGATGCGATGCGATGCGTGTGCTGATGATGCGATGCGATGCGTGTGCTGATG 80487
QY 116 aaattgctcaagtagacgccaagaatcaacggtlaagaacacccgtatggtattatg 175
Db 80486 ATGATGCTGTATGCTGCTGATGATGATGCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 80427
QY 176 caggttaacttgaccacaaatttggcgtagaacccgaattgtgtgtcagaagcca 235
Db 80426 ATGGTATGCGATGCGATGCGATGCGTGTGCTGATGATGCTGATGCGATGCGATGCGTGTG 80367
QY 236 aagaatttaatgcagcgctgctgtaaaagtgatgtgaagctcttggctctatg 295
Db 80366 ATGCTGATGACGATGCGATGCGTGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 80307
QY 296 gcacatacgctataactcaatcaat 321
Db 80306 GTAATGATGATTAATATATGATGATGT 80281
RESULT 8
AC005140/c 308399 bp DNA HTG 01-JAN-2000
LOCUS Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.
AC005140 AC005140.5 GI:6652490
VERSION HTG; HTGS_PHASE1.
KEYWORDS malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 308399)
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Rowley, D., Tamaki, T., Kurdi, O.B.,
TITLE Conway, A.B. and Davis, R.W.
JOURNAL Plasmodium falciparum 3D7 chromosome 12
REFERENCE 2 (bases 1 to 308399)
AUTHORS Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT On Jan 1, 2000 this sequence version replaced gi:5919273.
* NOTE: This is a 'working draft' sequence. It currently
```

```
*
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 66874: contig of 66874 bp in length
* 66875 67074: gap of unknown length
* 67075 177925: contig of 110851 bp in length
* 177926 178125: gap of unknown length
* 178126 225300: contig of 47175 bp in length
* 225301 225500: gap of unknown length
* 225501 241505: contig of 16005 bp in length
* 241506 241705: gap of unknown length
* 241706 247313: contig of 5608 bp in length
* 247314 257513: gap of unknown length
* 257514 257713: gap of unknown length
* 257714 277104: contig of 19391 bp in length
* 277105 277305: gap of unknown length
* 277305 308399: contig of 31095 bp in length.
*
* Location/Qualifiers
*   1. .308399
*     /organism="Plasmodium falciparum"
*     /db_xref="taxon:5833"
*     /chromosome="12"
*     /clone="3D7"
*
BASE COUNT 124664 a 29405 c 30157 g 122771 t 1402 others
ORIGIN
Query Match 7.8%; Score 42.2; DB 44; Length 308399;
Best Local Similarity 44.2%; Pred. No. 1;
Matches 173; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
QY 115 aaattgctcaagtagaagccaagaatcaacggtlaagaacacccgtatggtattat 174
Db 239260 AAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 239201
QY 175 gcaggttaacttgaccacaaatttggcgtagaacccgaattgtgtgtcagaagcc 234
Db 239200 GATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 239141
QY 235 aagaatttaatgcagcgctgctgtaaaagtgatgtgaagctcttggctctatg 294
Db 239140 AATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 239081
QY 295 ggcacatacgctataactcaatcaatcccaattttagccaaggcaattaggaatt 354
Db 239080 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 239021
QY 355 gctgaactaagtagatgcttgcacagcgctgaatgaactatactcaacaaagcgac 414
Db 239020 GATGACCAATGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 238961
QY 415 aaacccagccagcagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474
Db 238960 AATAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 238901
QY 475 gaagcaactaactatctatcagaagatg 505
Db 238900 CAAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 238870
RESULT 9
AF090946/c 1302 bp mRNA PRI 12-JAN-2000
LOCUS AF090946
DEFINITION Homo sapiens clone HQ0688.
AC005140 AF090946
AC005140 AF090946.1 GI:6690254
KEYWORDS FLI_CDNA.
SOURCE human.
```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1302)
AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
Zhou, S., Liu, M., and He, F.
TITLE Functional prediction of the coding sequences of 50 new genes
deduced by analysis of cDNA clones from human fetal liver
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1302)
AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
Zhou, S., Liu, M., and He, F.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1998) Dept. of Experimental Hematology, Beijing
Institute of Radiation Medicine, 27 Taiping RD, Beijing 100850,
P.R.China
FEATURES
source Location/Qualifiers
1. 1302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HQ0688"
/issue_type="liver"
/dev_stage="fetus"
BASE COUNT 324 a 217 c 179 g 415 t 167 others
ORIGIN
Query Match 7.7%; Score 41.6; DB 51; Length 1302;
Best Local Similarity 15.3%; Pred. No. 1.4;
Matches 38; Conservative 108; Mismatches 102; Indels 0; Gaps 0;
QY 24 gccagcattacgtctctgtattgagcagtgagtgctaacgctgcacatgacgataaggca 83
Db 793 GACGACGAGTGYKGRARRAMWKKRMMAAKKSYTTKSCMAARWYMTWARGSYNG 734
QY 84 tctcgtgagtcctcaacctatgttgccaaattgtgcaagtagcagccagaacaat 143
Db 733 GKTTKMRSMYTCAMMYRRAMWMTTTTMYKSMRSCCRMRKRAKKSSKSMWK 674
QY 144 caacgtaagaacacccgctatgtattatgcaggtataacttgacaaattgg 203
Db 673 GRMCWYMGRAARGRMWARGSYCYMRGYSKMRBARARWYMMMMRRMTTR 614
QY 204 cgtagaaccgaattgtgtgttcagaagccaaagaattatgcagcgtgagtcctgt 263
Db 613 AARGGRMYMTWMTWGSYKRAKRAKMAKRAAMWSPARRMTTWAARRGARAGKTYKR 554
QY 264 aaaagtg 271
Db 553 AAKMSWG 546
RESULT 10
AC016048/C
LOCUS AC016048 172633 bp DNA HTG 21-JAN-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-98G11, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC016048
VERSION AC016048.1 GI:6456151
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172633)
AUTHORS Bruno, D., Conn, L., Dela Rosa, M., Federspiel, N., Foreman, P.,
Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
Yu, S., and Davis, R.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172633)
AUTHORS Bruno, D., Conn, L., Dela Rosa, M., Federspiel, N., Foreman, P.,
Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,

COMMENT
JOURNAL
TITLE
AUTHORS

Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
Yu, S., and Davis, R.W.
Direct Submission
Submitted (20-NOV-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SPSSTDC
Web site: <http://sequence-www.stanford.edu/group/human/>
Contact: hum.info@sequence.stanford.edu
----- Project Information
Center project name: 705
Center clone name: RP11-98G11
----- Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-Primer; % of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 154547 bases at least Q40
Consensus quality: 162256 bases at least Q30
Consensus quality: 164167 bases at least Q20
Insert size: 177356; agarose-fp
Insert size: 171633; sum-of-contigs
Quality coverage: 5.2x in Q20 bases; agarose-fp
Quality coverage: 5.3x in Q20 bases; sum-of-contigs.
NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1248: contig of 1248 bp in length
* 1249 1298: gap of unknown length
* 1299 2671: contig of 1373 bp in length
* 2672 2721: gap of unknown length
* 2722 4038: contig of 1317 bp in length
* 4039 4089: gap of unknown length
* 4089 5263: contig of 1175 bp in length
* 5264 5313: gap of unknown length
* 5313 6770: contig of 1457 bp in length
* 6771 6820: gap of unknown length
* 6821 9900: contig of 3080 bp in length
* 9901 9950: gap of unknown length
* 9951 13563: contig of 3613 bp in length
* 13564 13613: gap of unknown length
* 13614 18849: contig of 5236 bp in length
* 18850 18899: gap of unknown length
* 18900 24187: contig of 5288 bp in length
* 24188 24337: gap of unknown length
* 24338 29472: contig of 5235 bp in length
* 29473 29522: gap of unknown length
* 29523 35549: contig of 6027 bp in length
* 35550 35599: gap of unknown length
* 35600 42925: contig of 7326 bp in length
* 42926 42975: gap of unknown length
* 42976 48502: contig of 5527 bp in length
* 48503 48552: gap of unknown length
* 48553 57258: contig of 8706 bp in length
* 57259 57308: gap of unknown length
* 57309 66203: contig of 8895 bp in length
* 66204 66253: gap of unknown length
* 66254 75336: contig of 8933 bp in length
* 75337 75286: gap of unknown length
* 75287 85275: contig of 9989 bp in length
* 85276 85325: gap of unknown length
* 85326 95182: contig of 9857 bp in length
* 95183 95232: gap of unknown length
* 95233 118006: contig of 22774 bp in length
* 118007 118056: gap of unknown length

* 118057 144296: contig of 26240 bp in length
* 144297 144346: gap of unknown length
* 144347 172633: contig of 28287 bp in length.

FEATURES
Source
1. 172633
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone_id="RPC1 human BAC library 11"

BASE COUNT 40832 a 46710 c 45544 g 38472 t 1075 others
ORIGIN

Query Match 7.6%; Score 41.4; DB 45; Length 172633;
Best Local Similarity 48.5%; Pred. No. 1.7;
Matches 114; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 77 atggcaattcgtcgatgctcaaccctatgtgtgcccagaattggtcaagtagagcgca 136
Db 76270 ATGGTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 76211
Qy 137 agcaaatcaacggtgaagaacacccgttatgtattatcgcagttataacttgaccaa 196
Db 76210 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 76151
Qy 197 atttgcgttagaaccgaattgtgtgttcagacgcgaagaatttaatcagagcggtga 256
Db 76150 ATGATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 76091
Qy 257 gtccctgaagaagtgatgtgaagtccttgggtgtgttaatgacacatcgcataa 311
Db 76090 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 76036

RESULT 11

SCU20616 1277 bp DNA PLN 22-JUL-1995
LOCUS SCU20616
DEFINITION Saccharomyces cerevisiae G4p2 gene, complete cds.
ACCESSION U20616.1 GI:710337
VERSION U20616.1 GI:710337
KEYWORDS
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.

REFERENCE

REFERENCE
AUTHORS Frantz,J.D. and Gilbert,W.
TITLE A yeast gene product, G4p2, with a specific affinity for quadruplex
nucleic acids
JOURNAL J. Biol. Chem. 270 (16), 9413-9419 (1995)
MEDLINE 95238458
REFERENCE 2 (bases 1 to 1277)
AUTHORS Frantz,J.D.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1995) J. Daniel Frantz, Biology, Harvard
University, Cambridge, MA 02138, USA
LOCATION/Qualifiers
1. 1277

FEATURES

source
/organism="Saccharomyces cerevisiae"
/strain="S288c"
/db_xref="taxon:4932"
/clone_id="pool of clones G4p2-1, -2, -3"
/clone_id="Yeast genomic DNA in lambda EMBL3a; Snyder et.
al. PNAS 83, p.730"
234. 1055
/note="Similar to yeast suppressor protein MPT4,
Swiss-Prot Accession Number P39015"
/product="G4p2"
/codon_start=1
/protein_id="AA070169.1"
/db_xref="GI:710338"
/translation="MSNPFDLLGNVEDADVYVLPKEIVKSNSTSKKADVPPPSADP

SKARNRPRSGNEGATRDKTAGRRNRNSKDVDTATTKSNFRATDRSRGKTDTP
KKYKNCWGDKKRELSAEKQADAAAEIADAEADADGAPKPAQSLSDYLYNQAN
KQNKVPEAKEVEILDAGRIETAEKEAYPATKAVNSKOLKTEILEFDIVESNT
RKNFGRDNNRNNRNNRGRGRKARNANANVSNATYQKNRIDVSNLPSLA"
BASE COUNT 415 a 262 c 250 g 350 t
ORIGIN

Query Match 7.5%; Score 40.8; DB 7; Length 1277;
Best Local Similarity 44.7%; Pred. No. 2.3;
Matches 159; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Qy 91 gatgctaacccctatgttggcgccaaattggtcagtagagcgcaagcaatcaacggt 150
Db 597 GAAGCTCAAGCCGATGCTGCTGCTGAAATTTGCTGAAGACGCTGCAGAACTGAAGACGCT 656
Qy 151 aagaacaccgcttatgattatcagttataacttgcacaaatttggcgtagaa 210
Db 657 GGTAGCCAAAGACCGCTCAATTTGCTTTGCAAGACTACTTAACCAACACTTAACAC 716
Qy 211 ccggaattgtgtgttcagacgcgaagaatttaatgcagcggtgagtcgtlaaaggt 270
Db 717 CAGTTCACAAAGGCTCCAGAAAGCTTAAGAGGTTGAATTAGACGCTGGAAGATTGAAC 776
Qy 271 gatgtgaagtccttgggtgtgtgtatgacacatcgtctaatcattcaatccccatt 330
Db 777 GCTGAAAAGGAGCTTTACGTTCCACCAACCAAGGTCAAGACGCTCAAACTTAAGCAATTG 836
Qy 331 tatgcagagggcaattagtcgcatgtcgaagtaagtagatgtagtcagcgcttagca 390
Db 837 AAGACCAAGACTACTTGGAAATTTGATGCCACTTTGTTGAATCTAAGCTGAAAGAAC 896
Qy 391 actacatactcaaaaaagcgacaaacccagcctagaagcggtgtgtgtgtg 446
Db 897 TTCGGTGACAGAAACACACACAGCAAAACACTTCAACCAACCGCTGGTGTAG 952

RESULT 12

AE001127/c 10976 bp DNA BCT 15-DEC-1997
LOCUS AE001127/c
DEFINITION Borrelia burgdorferi (section 13 of 70) of the complete genome.
ACCESSION AE001127 AE000783
VERSION AE001127.1 GI:2688047
KEYWORDS
SOURCE
ORGANISM Lyme disease spirochete.
Borrelia burgdorferi
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
burgdorferi group.
1 (bases 1 to 10976)

REFERENCE

REFERENCE
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
Gwin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,
Gocayne,J.D., Weidman,J., Uterback,T., Wathey,L., McDonald,L.,
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
TITLE Genomic sequence of a Lyme disease spirochete, Borrelia
burgdorferi
JOURNAL Nature 390 (6660), 580-586 (1997)
MEDLINE 98065943
REFERENCE 2 (bases 1 to 10976)
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
Gwin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,
Gocayne,J.D., Weidman,J., Uterback,T., Wathey,L., McDonald,L.,
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

```
FEATURES
source
    location/Qualifiers
        1..10976
        /organism="Borrelia burgdorferi"
        /db_xref="taxon:139"
gene
    107..823
    /gene="BB0158"
    107..823
    /gene="BB0158"
    /note="similar to PID:1063419 GB:AB000790 percent
    identity: 37.28; identified by sequence similarity;
    putative"
    /codon_start=1
    /transl_table=11
    /product="anti-gen, S2, putative"
    /protein_id="AAC66550.1"
    /db_xref="GI:2688051"
    /translation="MVFPRYKHLLEIMLPMMLSCAFPKKPOSHODSNTGKSIDEX
    IHLISGKTSNKKLPIIINSNDYTWIKTKAMTILSGDGKEIPFKNKFGSIYISPKM
    DKRISYASLILFETTKNGDDEYIEDKFTAGSTLEKNSLAVENSQEGRYTA
    IPEGLMSDEIKNAKRLTYKNGHWMYMLADLVKMKLVOETKIKYISLNSKLIIEFLK
    EVLKENSILKDIAGLDFEDI"
    847..1521
    /gene="BB0159"
    847..1521
    /gene="BB0159"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BB0159"
    /protein_id="AAC66555.1"
    /db_xref="GI:2688056"
    /translation="MSIKKFIILTLIILSLAKNSPSENEINIPENENYIKENIKTEIK
    KIKOFLASVDVAISOPIELADNGEPIKELGSIYSFIVNFSKIGSALISPDLS
    NEASKKTYIKLEFLSPDKGNFTINOLSSITSGQSKKELADAVSPGLPRESISKT
    IAEYKKNMYIILAIVENNINKETEKYERINPKIYNDQKRLHFKSNQIKKE
    PIPIE"
    1569..2687
    /gene="BB0160"
    1569..2687
    /gene="BB0160"
    /note="similar to SP:P29743 PID:396388 GB:U00096
    PID:1790487 percent identity: 29.89; identified by
    sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="alanine racemase (alt)"
    /protein_id="AAC66549.1"
    /db_xref="GI:2688050"
    /translation="MYNNKTMGSNKTIIINLNLEHNLINKIGEKIYATLKGDA
    YGHGLINFEKFLSKNINYPFGISNEDAKTLKIDKNIKILMYIKVKKKEIKNLKFE
    LPEVSDPEYFLIEKCALOKNKIKVHLKIDIGNNRGKIDDLALETATYONSKPO
    ELEGICSHLPSIENKTKOKOJEOFLFPLETLKOKNHKPKVHISNGSLIYNKQPO
    ENMVRPGILIGYCOSLKNKKPALNFKPYLSFSVITIKVKKGEKISYSQIPAKE
    DMKILIPIGFDGIPONISNDFYFLINKKCKIKGVCOMLTIIVEIPDLKVKGSK
    VEIYSEKLSIDEMSEKFSRSHYELLCNIGKYEKKRYLD"
    complement(2679..4343)
    /gene="BB0161"
    complement(2679..4343)
    /gene="BB0161"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BB0161"
    /protein_id="AAC66554.1"
    /db_xref="GI:2688053"
    /translation="MIDLLYNNKIYONRYVLSNLFMLPIIFISAPLSFEKVEYVFAD
    YSNYSIRIGKEITKESEEEYGFYIDRNMLLIVYIGKLSILMSPFNAQIKI
    ESKRSESRIFLNGLAKNNNGVYIEHLEWSDGRKININENYNSNEMVRDKYVSY
    YHILYDNRESVNRESGIOIKDLNVNFTKISYDHRKPKATLYIDENKIYRSK
    NGTGFRLTYDANHLIKEEYLDKSAIVNSPVSATKIYHVDVGKVEILNVTYINN
    NLTPENNVAIYREYIFKDKCYKKEYNNDNNMLTISQNGYAMKKTIFYIYONNEKR
```

```
gene
    4357..4524
    /gene="BB0162"
    4357..4524
    /gene="BB0162"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BB0162"
    /protein_id="AAC66553.1"
    /db_xref="GI:2688054"
    /translation="MAKISKNNQVRSKELLISRMGKIIIMSKFENGRIKHYAECOTS
    KNTAKRPRLP"
    complement(4551..6299)
    /gene="BB0163"
    complement(4551..6299)
    /gene="BB0163"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BB0163"
    /protein_id="AAC66552.1"
    /db_xref="GI:2688053"
    /translation="MKKLEPIFYMSKSEVDLILSOLGSISSDPQKSLTKLTSYKKEAI
    DTSFIEHSVKIRLKGELDFLFRKLVEKRGVATNRATGDELPBQIILTDASYS
    DEYYSIDLEFLKTYIGIELFSSLLTVKNIESSLVKNFETIADSDINKYFEFTKS
    QKEILVMSKLIIPSSRALYFEYPAKIFFNASNNVADVVRIRGLKILIDYEKLE
    EGNKIGASIKLSIGIEIKDEVLGKKGITFSKNYEVLEPINLPFKNEKDLBEKAI
    EAVKIDESIKALEKTELEKKSIDNEEPFGILKVVSEGLNDPENLIVANMYPRSKAI
    HDILKIQVKNFPLIKPRAPFELEEKYAKLNSLRASITRTIISFGDPSSEIF
    TEEGLEVLSRLVRSYITMNEYCLKNNKFLDLIVSHYRITCTQSVQVHLKDLIVEA
    AAYFKNSNELPLRYIKLIDLSILNAAYKDLSPKRIYVLTGTHQSYKEALSRIDF
    RAGLKNKSGSLFSDSNERYKQLEIRKOREEIKEQVRNSFPKGAAPQAVASKSYTKQ
    EODEAMVKEFSDRIOKK"
    complement(6380..7393)
    /gene="BB0164"
    complement(6380..7393)
    /gene="BB0164"
    /note="similar to GB:L77117 SP:057556 PID:1590872 percent
    identity: 34.47; identified by sequence similarity;
    putative"
    /codon_start=1
    /transl_table=11
    /product="Na+/Ca+ exchange protein, putative"
    /protein_id="AAC66548.1"
    /db_xref="GI:2688049"
    /translation="MPLFVKEIYIOHLIOFFVVGRCIFILYIGNLLKSSGCIATYL
    KYPTILGVTIYSETSABELTSLVAARVKGNEIVSVNSVIGSNTIIMLALPLAGFF
    LKIKRDLFRKLTSEFMFLMLLILLLSDFDSYFEPVYPRYRFSFSLILFLSYLL
    LFYKEKEKAYSLSENSFOGASNLNOSFNEIFPNLSFVISMVFLVGLSGLVADGLYI
    ANNVNVESEKILGIIILVAFGVSPLVYSLFAIIRKESDIAGNTEGSIINFGIYLA
    SSSFPRLIDQIYILDSIMVETFLVFLVYKFGVGRGSLIFLLIYFINMLFM
    NPY"
    complement(7394..9238)
    /gene="BB0165"
    complement(7394..9238)
    /gene="BB0165"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BB0165"
    /protein_id="AAC66551.1"
    /db_xref="GI:2688052"
    /translation="MYMKIFITWVYIFPFSVKKVYSIYSLNDEEFPKYSJLFYHKG
    LSKNNGKITKQVNGVINSRWYPPYKLPVPSITSIYEDVYSSSSFLTSSNLLVYSYD
    YSKNRLKLVGIDKFNAGVITSSASQDQKRIATGIGHIYLSVNGAISPKNNRL
    IPOILVAGAYDIISAIEFSKBEETNNLVFSSGVYDIFLISQSGFIKIKISPPFKQI
```


IRILDSSKNVEKILVPTYNHFSYINGOWFIKLSLQDQDFEKSORMLAKNKG
SIYLATYILRNKKAVIDERKFIKDSGMAAVYIDFDDGNDLTLTYSKLSLPNKLRSVRN
FLDPEYILKKAKELGIYIARCYVPRKDKLYVIDFQKIALNKNKRNKRAHLIKRVD
SGLYAVYVEHWIDIFSPATMEYNISIKAEIOSFVDELOFYIIFPSGVLAIYS
MKRYEMOPVDALIESFLIMAREQLYVISVDIYGYNGWEPNLSIGNISMLSDYVIS
PMFYSHYTDDELPSNFYTKRAYRIYEGSDRALAFSLDGVIIAPYVOAFILGKERL
VDDETLYLKEFKLKEIKESFSGCSFLMNASNVYMIKSLKEYLDSF"
complement(9358. .10878)
/gene="BB0166"
complement(9358. .10878)
/note="similar to PID:1651873 percent identity: 38.63;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="4-alpha-glucanotransferase (malQ)"
/protein_id="AAC6547.1"
/db_xref="GI:2688048"
/translation="MKYKKRINLNLKRRSGILNMISSLPKSGIGDLGKAKYKFLDF
LFASQSYWQMFAYSPIDFTSRSPYSIFSAFAGNYYIDLELDFIDSLKLENE

Query Match
Best Local Similarity 7.4%; Score 40.2; DB 2; Length 10976;
Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 163 tatgtattatgcaggtataacttgcacaaatttggcgtagaaccgaattgtc 222
Db 5750 TTTGTTTAAAAAATATTTTATATATTTGCTTCAATAATGTTTAAAGATGTTTCGTT 5691

Qy 223 ggttcagacgcaagaattatgcagcgtagcctgtaaaagtgatgtagaagtc 282
Db 5690 ACAATATAAGGATTAAAGCTTATGATTTAGACAAAACCTTTAGAAAGGTTAAGCAATATT 5631

Qy 283 ttgtgtctatgacacatcgcataacttcaatcaatccatttatgcgaagggc 342
Db 5630 ATTGGTGCTATTTCCAAAGATTCCTCGTATATGAAATCAAGATGAAGTTCTTGCGC 5571

Qy 343 aaattagcattgctgaagactaaagtatgctc 375
Db 5570 AAAAAAGAAATTACTTTTCCAAAAAATTAATTTT 5538

RESULT 13
128501 I28501 846 bp DNA PAT 06-FEB-1997
LOCUS
DEFINITION Sequence 21 from patent US 5571718.
ACCESSION I28501
VERSION I28501.1 GI:1819277
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN

1 (bases 1 to 846)
Dunn,J.J. and Barbour,A.G.
Cloning and expression of soluble truncated variants of Borrelia
OspA, OspB and Vmp7
Patent: US 5571718-A 21 05-NOV-1996;
Location/Qualifiers
1. 846
/organism="unknown"
367 a 134 c 150 g 195 t

Query Match
Best Local Similarity 7.3%; Score 39.4; DB 5; Length 846;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 33 agctcttcgtatttggcgaatgagctgaagctgcacatcgaatctgctga 92
Db 117 AGACTCAGTGTCTTTGTTAATGTAATAAATTTTGTAAAGCAAGAAAAAATAGCTC 176

Qy 93 tgctcaacctatgtctggtgcacaaattgctcaagtagacgcaagaatcaacggttaa 152

CGGCAATAATGATTTAAAGCAACAAFTGATGAGTTGAACCTTAAAGCACTCCGATA 236
Qy 177
Db 153 gaacacgcttatgattatgcaggtataacttgcacaaatttggcgtagaacc 212
Db 237 AACAATAGTGTTCGGAACCTTGAAGCTTCAAGCCTGACAAAGTAAGTAATAATTAAAC 296

Qy 213 cgaatttggcttcagcgcaagaattatgca 249
Db 297 AGTTCTGCTGATTTAAACACAGTAACCTTGAAGCA 333

RESULT 14
AF077664
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS

AF077664 873 bp DNA BCT 03-NOV-1998
Borrelia burgdorferi arthropod-specific outer surface protein B
(ospB) gene, partial cds.
AF077664
AF077664.1 GI:3386498
Lyme disease spirochete.
Borrelia burgdorferi
Bacteria: Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
burgdorferi group.
1 (bases 1 to 873)
Ryan,J.R., Levine,J.F., Apperson,C.S., Lubke,L., Wirtz,R.A.,
Spears,P.A. and Orndorff,P.E.
An experimental chain of infection reveals that distinct Borrelia
burgdorferi populations are selected in arthropod and mammalian
hosts
Mol. Microbiol. 30 (2), 365-379 (1998)
99009365
2 (bases 1 to 873)
Ryan,J.R., Levine,J.F., Apperson,C.S., Lubke,L., Wirtz,R.A.,
Spears,P.A. and Orndorff,P.E.
Direct Submission
Submitted (13-JUN-1998) Department of Microbiology, Parasitology
and Pathology, North Carolina State University, College of
Veterinary Medicine, 4700 Hillsborough Street, Raleigh, NC 27606,
USA
Location/Qualifiers
1. 873
/organism="Borrelia burgdorferi"
/strain="P235.1"
/db_xref="taxon:139"
<1. .>873
/gene="ospB"
<1. .>873
/gene="ospB"
/note="ospB"
/codon_start=1
/transl_table=11
/product="arthropod-specific outer surface protein B"
/protein_id="AAC69850.1"
/db_xref="GI:3386499"
/translation="IGFALALALIGCAQKGAISGQKENDLNLEDSSKSHQAKOD
LPAVEDSVSLFNGNKIFVSKENSSGKDLRATIDQVELKGTSDNNGSGTLEGSKP
DKSKVLIYASDLNVTLEAFDASNOKISSKTKKQSGITEETLKANKLDSKRLTRSN
GTTLEISQITDADNATKAVETLKNSIKLSLVGKTKTVEIKGTVTLKRETEKDKV
KVIINDTASNNKTKGKWDSTSTLTISASSKTKDLVFLDGIITVQVINTACTSLG
SASEIKNLSELKNAI"
370 a 137 c 159 g 207 t

Query Match
Best Local Similarity 7.3%; Score 39.4; DB 2; Length 873;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 33 agctcttcgtatttggcgaatgagctgaagctgcacatcgaatctgctga 92
Db 150 AGACTCAGTGTCTTTGTTAATGTAATAAATTTTGTAAAGCAAGAAAAAATAGCTC 209

Accession	Feature	Start	End	Score	Value	Description
Db	174	TGGCAATATGATTTAAGGACACAAATGTGATCGGTTGAACCTTAAGGAACCTCCGATAA	233			
Qy	153	gaacacccgtatcgtgattatcgaagctataacttgacattgacaaatttgcgtagacc	212			
Db	234	AAACAATGGTCTCGAAGACCCCTTGACGTTCAAGCCTGCACAAAGTAAGTAAATTAATTAAC	293			
Qy	213	ggaatttgttggttcagacgcgcgaagaattaatga	249			
Db	294	AGTTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA	330			
RESULT	2					
Id	V74423	standard; DNA; 18613 BP.				
AC	V74423;					
DT	16-MAR-1999	(first entry)				
DE	Staphylococcus aureus contig SRQ ID #112.					
KW	Computer readable medium; vaccine; S. aureus infection; immunodetection;					
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;					
KW	skin infection; surgical wound infection; scalded skin syndrome;					
KW	toxic shock syndrome; ds.					
OS	Staphylococcus aureus.					
Key	Location/Qualifiers					
FT	misc_feature	901..960				
FT		/tag= a				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	2701..2760				
FT		/tag= b				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	4501..4560				
FT		/tag= c				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	6301..6360				
FT		/tag= d				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	8101..8160				
FT		/tag= e				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	9901..9960				
FT		/tag= f				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	11701..11760				
FT		/tag= g				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	13501..13560				
FT		/tag= h				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				
FT	misc_feature	15301..15360				
FT		/tag= i				
FT		"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"				

```

FT      /*tag= 1      /note= "these bases represent a line of missing text in
FT      /note= "the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
FT      misc_feature      17101.      17160
FT      /*tag=      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
PD      BP-786519-A2.
PD      30-JUL-1997.
PR      07-JAN-1997; 100117.
PR      05-JAN-1996; US-009861.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI      Rosen CA;
PI      WPI; 97-374922/35.
PI      Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PI      stored on computer readable medium and used in the production of
PI      anti-S. aureus vaccines
PS      Claim 1, Page 641-651; 3271pp; English.
CC      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC      of the invention. The DNA sequences are recorded on a computer readable
CC      medium, preferably selected from a floppy or hard disk, random access
CC      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC      the S. aureus DNA sequences allows putative functions to be assigned so
CC      that protein-encoding or regulatory regions of commercial, therapeutic or
CC      industrial importance can be obtained. Specifically, sequences which are
CC      likely to encode antigens have been identified and these polypeptides can
CC      be used in a vaccine composition against S. aureus infection. The
CC      polypeptides can also be used in a kit for the immunodetection of
CC      S. aureus in a sample. S. aureus is implicated in numerous human diseases,
CC      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC      skin and surgical wound infections, scalded skin syndrome, toxic shock
CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
CC      for recombinant production of the polypeptides. The new DNA sequences
CC      (and their fragments) are useful as primers or probes for isolating
CC      homologues of any of the S. aureus DNA sequences contained on the
CC      computer readable medium.
SQ      Sequence 18613 BP; 5990 A; 2753 C; 3477 G; 5782 T;

Query Match      7.4%; Score 40.4; DB 1; Length 18613;
Best Local Similarity 51.7%; Pred. No. 0.017;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

OY      91      gatgtcaaccatattgtgtgcacaaatgtgtcaagtagacgcgcgaagcaatacaagcgt 150
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      5362      GATGTCATACGATTCATCATGAGACATGGGTACCTTTATACCAACACGTGCAATTG 5421

OY      151      aagaacaccgcgtattgtattattatcagcgtataactttgacccaanaatttggcgtagaa 210
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      5422      GAAACAACAAAGCGTCGATCTTTATGAAGGTAAATGAAAGTGAGTAAGCA 5481

OY      211      ccgaattgtgtgttcagcgcgaagaatttaatcagcgcgtgagtcctcgttaagaag 268
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      5482      AAGCTTGATTGATGACGAGTTGTAATGATCACTGAACACGATTGAAACCGGAACAAG 5539

RESULT      3
Continuation (8 of 10) of X20248 from base 700001 (Borrelia burgdorferi polynucleotide
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP      Fragment Name      Begin      End
WP      X20248_00      1      110000
WP      X20248_01      100001      210000
WP      X20248_02      200001      310000
WP      X20248_03      300001      410000
WP      X20248_04      400001      510000
WP      X20248_05      600001      710000
WP      X20248_06      800001      910000

```

WP X20248_07 700001 810000
 WP X20248_08 800001 910000
 WP X20248_09 900001 910715

Query Match 7.4%; Score 40.2; DB 1; Length 110000;

Best Local Similarity 49.3%; Pred. No. 0.043;

Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 163 tatgtattatgcaggttataactttgacaaaatttggcgtagaaccgaaattgtt 222
 Db 46445 TTTGCTTAAAAAATATTTTAAATATGCTCTTAATATGTTAAAAAGATGTTTCGT 46504

QY 223 ggttcacagcccaagaatttaacgcagcggtgtgtctgtgaaagtgtgaagctc 282
 Db 46505 AGAATTAAGGATTTAAAGCTTATGATTAAGAAACCTTTAAGAGGGTAAAGAAATTT 46564

QY 283 ttgtgtctatggcacatcatcactcatcaataaccattttatgcgaagggc 342
 Db 46565 ATTGCTGCTATTTCACAGATTCTCTGCTGATTAATGAAACAGATGAAGTCTTGCC 46624

QY 343 aaattagcattgtcctaagaactaagtagatgt 375
 Db 46625 AAAAAAGGAATTACTTTTCAAAAAATATATTT 46657

RESULT 4
 T43302
 ID T43302 standard; DNA; 846 BP.

AC T43302;
 DT 10-FEB-1997 (first entry)

DE Ospb soluble variant coding sequence.

KW OSPA; Ospb; outer surface protein A; Borrelia; variable major protein 7;
 KW Borrelia hermslii; Vmp7; surface lipoprotein; spirochete; human; antigen;
 KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;
 KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;
 KW ds.

OS Borrelia burgdorferi.
 PN US5571718-A.
 PD 05-NOV-1996.

PF 08-SEP-1992; 941523.
 PR 21-DEC-1990; US-632072.
 PR 08-SEP-1992; US-941523.

PA (ASU-) ASSOC UNIVERSITIES INC.
 PI Barbour AG, Dunn JJ;
 DR WPI; 96-505409/50.
 DR P-PSDB; W08096.

PT Soluble recombinant forms of Borrelia lipo:proteins - useful for
 vaccine prodn. for treatment of Lyme disease

PS Claim 4; Column 43-44; 49pp; English.

CC T43301-T43303 represent coding sequences for soluble recombinant Borrelia
 proteins of the invention. This sequence represents the coding sequence
 for the soluble recombinant variant of the Borrelia burgdorferi outer

CC surface protein B (Ospb). Ospb is a surface lipoprotein of the B.
 CC burgdorferi spirochete. Borrelia spirochetes are responsible for a
 variety of human disorders including Lyme borreliosis, and relapsing
 fevers. The spirochete is transmitted to humans and animals through the

CC bite of a tick, and can cause serious dermatological, arthritic,
 CC neurological and other pathogenic disorders in an infected host. This
 CC sequence is used to create recombinant host cells, and the encoded Ospb
 CC protein can be isolated from the cytosol of one of these cells without
 CC the use of detergent. The encoded recombinant proteins can be used as

CC antigens for the production of vaccines against Lyme disease. The
 CC recombinant proteins can also be used in immunoassays and other
 CC diagnostic screening methods to detect the presence of antibodies against
 CC Borrelia lipoproteins in the sera of infected patients.

CC Sequence 846 BP; 367 A; 134 C; 150 G; 195 T;
 SQ

Query Match 7.3%; Score 39.4; DB 1; Length 846;
 Best Local Similarity 48.8%; Pred. No. 0.0087;
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 33 agcttcctgtattggcgcagtagtgcgaacgctgcacacagcagctatctgcctga 92
 Db 117 AGACTCAGTGTCTTTGTTTAAATGCTAATTAATTTTGTAAAGCAAGAAAAATAGCTC 176

QY 93 tgcacaacctatgttggtccaaatgtgcgaagtagacgcgaacaaatcaacggtga 152
 Db 177 CGGCAAAATATGATTTAAAGACAAATTAAGATGAGTTCAGTTAAAGAACTTCGATTA 236

QY 153 gaacacgcctatgtatctatgcaggtataacttgcacaaaatttggcgtagaac 212
 Db 237 AAACAAATGTTCTGTGAACCTTGAAGGTTCAAAAGCCTGACAAAGATAAGTAATTAAC 296

QY 213 cgaatttggttcagcagcgaagaatttaatga 249
 Db 297 AGTTTCTGTATTTAAACACAGTAACCTTGAAGCA 333

RESULT 5
 ID 090713 standard; DNA; 891 BP.

AC 090713;
 DT 31-JUL-1996 (first entry)

DE B. burgdorferi strain B31 outer surface protein B (Ospb-B31) DNA.
 KW Strain B31; outer surface protein; Ospb; antigenic domain;
 KW chimeric protein; treatment; diagnosis; infection; vaccine;
 KW Lyme borreliosis; immunodiagnostic assay; antibody;
 KW T-cell reactivity; chimeric; ds.

OS Borrelia burgdorferi.
 FH Key Location/Qualifiers
 FT cds 1..891
 FT 1..891

PN W09512676-A1.
 PD 11-MAY-1995.
 PR 27-OCT-1994; U12352.
 PR 01-NOV-1993; US-148191.
 PR 29-APR-1994; US-235836.

PA (ASU-) ASSOC UNIVERSITIES INC.
 PI Dunn JJ, Luft BJ;
 DR WPI; 95-215034/28.
 DR P-PSDB; R75726.

PT Chimeric protein comprising 2 or more antigenic Borrelia
 PT polypeptides) - useful in a vaccine against Lyme borreliosis and in
 PT immuno:diagnostic assays

PS Example 1; Fig 11; 20pp; English.

CC The present sequence encodes the B. burgdorferi strain B31, outer
 CC surface protein B (Ospb-B31). Using chemical or enzymatic methods,
 CC peptide fragments of Ospb-B31 were prep., and analysed by western

CC blot to assess their ability to bind different anti-Ospb monoclonal
 CC antibodies. The information obtd. was used to locate antigenic
 CC domains in Ospb-B31, the epitopes of which were mapped with the

CC aid of site directed mutagenesis. Identical analyses were performed
 CC on a selection of Osp purified from a variety of B. burgdorferi
 CC strains, the results from which were utilised in the prep. of a
 CC pool of antigenic Borrelia polypeptides, and corresponding

CC polynucleotides. Chimeric proteins comprising 2 or more antigenic
 CC Borrelia polypeptides, that do not naturally occur in the same
 CC protein, can be used in the treatment and diagnosis of Borrelia
 CC infections, i.e. as a vaccine against Lyme borreliosis, in
 CC immunodiagnostic assays to detect anti-Borrelia antibodies or to
 CC measure T-cell reactivity.

CC Sequence 891 BP; 380 A; 137 C; 160 G; 214 T;
 SQ

Query Match 7.3%; Score 39.4; DB 1; Length 891;
 Best Local Similarity 48.8%; Pred. No. 0.0089;
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 33 agcttcctgtattggcgcagtagtgcgaacgctgcacacagcagctatctgcctga 92
 Db 162 AGACTCAGTGTCTTTGTTTAAATGCTAATTAATTTTGTAAAGCAAGAAAAATAGCTC 221

QY 93 tgcacaacctatgttggtccaaatgtgcgaagtagacgcgaacaaatcaacggtga 152
 Db 237 AAACAAATGTTCTGTGAACCTTGAAGGTTCAAAAGCCTGACAAAGATAAGTAATTAAC 296

Accession	Sequence	Position
Db	222 CGCGCAATATGATTTAAAGACCAACATTTGATGAGTTGCACTTAAAGCAACTTCGATTA	281
Qy	153 gaacacgcctatggtgcatatgacaggtatataacttggacaaaatttggsgtgaac	212
Db	282 AAMCAATGGTTCTGGAAACCTTTAAAGTTCAAAAGCTGACAAAGATTAAGTAAATTAAC	341
Qy	213 cgaattgtgtgtcgaagcgcaagaattatgca	249
Db	342 AGTTTCTGCTGATTTAAACACAGTAACTTAGAAGCA	378

6 RESULT

ID	TA33315; standard; DNA; 891 BP.
DT	10-FEB-1997 (first entry)
DE	OspB variant #1 coding sequence.
KW	OspA; OspB; outer surface protein A; Borrelia; variable major protein 7;
KW	Borrelia hemissi; Wm7; surface lipoprotein; spirochete; human; antigen;
KW	Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;
KW	athritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;
OS	ds.
PN	Synthetic.
PD	US5571718-A.
PF	05-NOV-1996.
PR	08-SEP-1992; 941523.
PR	21-DEC-1990; US-632072.
PR	08-SEP-1992; US-941523.
PA	(ASU-) ASSOC UNIVERSITIES INC.
PI	Barbour AG, Dunn JJ;
PI	WPI; 96-505409/750.
PT	Soluble recombinant forms of Borrelia lipoproteins - useful for
PT	vaccine prodn. for treatment of Lyme disease
PS	Example 8; Column 39-42; 49pp; English.
CC	This sequence represents the coding sequence for a soluble recombinant
CC	Borrelia outer surface protein B (OspB) variant proteins of the
CC	invention. OspB is a surface lipoprotein of the B. burgdorferi
CC	spirochete. Borrelia spirochetes are responsible for a variety of human
CC	disorders including Lyme borreliosis, and relapsing fevers. The
CC	spirochete is transmitted to humans and animals through the bite of a
CC	tick, and can cause serious dermatological, arthritic, neurological and
CC	other pathogenic disorders in an infected host. This sequence is used to
CC	create recombinant host cells, and the encoded OspB protein can be
CC	isolated from the cytosol of one of these cells without the use of
CC	detergent. The encoded recombinant proteins can be used as antigens for
CC	the production of vaccines against Lyme disease. The recombinant
CC	proteins can also be used in immunoassays and other diagnostic screening
CC	methods to detect the presence of antibodies against Borrelia
CC	lipoproteins in the sera of infected patients.
CC	Sequence 891 BP; 380 A; 137 C; 160 G; 214 T;

Query Match	7.3%	Score 39.4	DB 1	Length 891
Best Local Similarity	48.8%	Pred. No. 0.0089		
Matches 106, Conservative	0	Mismatches 111	Indels 0	Gaps 0

[illegible]

QY 213 cgaatttgttgcagacgcccaagaatttaatgca 249
 | | | | | | | | | | | | | | | | | |
 Db 342 AGTTTCTGCTAATTTAAACACAGTAACTTGAAGA 378

RESULT 7

ID	Q90737	standard; DNA; 1141 BP.
AC	000737.	

AC Q90737; 30 1000

DE B31 outer surface protein (Osp-A)/antigen P41 (140-234) fusion gene.

KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;
KW immunodiagnostic assay; antibody; T-cell reactivity.

KW outer surface protein; Osp-A; antigen P41; fusion; ds.

FD	Key	Location/Qualifiers
FD	Key	Location/Qualifiers

```

E1 mal_peptide 1. :1141
E1 /*tag= a

```

PN	WO9512676-A1
PD	11-MAY-1995

PF 27-OCT-1994; U12352

PR 29-APR-1994; US-235836

PA (ASUY-) ASSOC UNIVERSITIES INC.
PI Dunn JT. Luft BJ:

DR WPI; 95-215034/28.

PT Chimeric protein comprising 2 or more antigenic *Borrelia*

PT immuno:diagnostic assays

Claim 43; Fig. 35; 200pp; English.
The present sequence encodes a fi

CC surface protein A (Osp-A) and antigen P41 (140-234) sequences,

CC peptide fragments of the parent proteins were prepd., and analysed

CC anti-Osp-A/P41 monoclonal antibodies. The information obtd. was used

CC were mapped with the aid of site directed mutagenesis. Identical
CC to locate antigenic domains in the proteins, the epitopes of which

CC analyses were performed on a selection of antigens purified from a variety of Burkholderia strains. The results are presented in Table 1.

CC utilised in the prepn. of a pool of antigenic *Borrelia* polypeptides,

above protein) comprising 2 or more antigenic Borrelia polypeptides,

CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine

CC against Lyme borreliosis, in immunodiagnostic assays to detect anti-Borrelia antibodies or to measure B-cell reactivity.

Sequence 1141 BP; 449 A; 206 C; 221 G; 265 T;

Query Match	7.38;	Score 39.4;	DB 1;	Length 1141;
Best Local Similarity	48.88;	Pred NO	0.0099;	

Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

114 AGACTCAGTGTCTTTTAAATGCTAATAAAATTTTGTAAAGCAAGAAAAATAGCTC 173

174 CGGCAAAATGATTAGAGCAACAATTGATCAGGTGAACCTTAAAGGAACCTTCGATTA 233

Dy 153 gaacaccgctatgcgtattcatgcaagttatacaactttaaaccaaatcttcgcgtagaacc 212
||| | ||| | ||| ||| | | |
Db 234 AAACAATGCTTCTGGAAACCTTTGAAGGTTCAAAAGCCTGACAAAGATTAAACTAAAAATTAAAC 293

```

Oy      213 cgaatttgtgttcagacgcgaagaatttaatgca 249
          | | | | | | | | | | | | | |
Db      294 AGTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 330

```

RESULT	8
Q90735	
ID	Q90735 standard; DNA; 1180 BP.

DT 30-JUL-1996 (first entry)

PD 11-MAY-1995. 012352.
 PF 27-OCT-1994; 012352.
 PR 01-NOV-1993; 05-148191.
 PR 29-APR-1994; US-235836.
 PA (ASUY-) ASSOC UNIVERSITIES INC.
 PI Dunn JF, Luft BJ.
 PI WPI: 95-215034/28.
 DR P-PSDB: R75743.
 PT Chimeric protein comprising 2 or more antigenic Borrelia
 PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in
 PT immuno-diagnostic assays
 PS Claim 43: Fig 34: 200pp: English.
 CC The present sequence encodes a fusion protein comprising outer
 CC surface protein A (OspA) and antigen P41 (122-293) sequences, from
 CC the B. burgdorferi strain B31. Using chemical or enzymatic methods,
 CC peptide fragments of the parent proteins were prep., and analysed
 CC by western blot to assess their ability to bind different
 CC anti-OspA/P41 monoclonal antibodies. The information obtd. was used
 CC to locate antigenic domains in the proteins, the epitopes of which
 CC were mapped with the aid of site directed mutagenesis. Identical
 CC analyses were performed on a selection of antigens purified from a
 CC variety of B. burgdorferi strains, the results from which were
 CC utilised in the prep. of a pool of antigenic Borrelia polypeptides,
 CC and corresponding polynucleotides. Fusion proteins (i.e. the
 CC above protein) comprising 2 or more antigenic Borrelia polypeptides,
 CC that do not naturally occur in the same protein, can be used in the
 CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine
 CC against Lyme borreliosis, in immunodiagnostic assays to detect
 CC anti-Borrelia antibodies or to measure T-cell reactivity.
 SQ Sequence 1363 BP; 538 A; 241 C; 252 G; 332 T;

Query Match 7.3%; Score 39.4; DB 1; Length 1363;
 Best Local Similarity 48.8%; Pred. No. 0.011;
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 33 agctcttcgtatgtgagatgagtgctaacgtccatcaactatggaattctgctga 92
 DB 114 AGACTCAGTGTCTTGTATTAAAGTAATTAATTTTGTAAACCAAGAAAATAAGCTC 173
 OY 93 tgcacacctatgtgtgccaatgtgcaagtacagtcgaagcgcaacaatcaacgtaa 152
 DB 174 CGGCAAAATATGATTTAAAGCAACAATTCATCAGTTGAACCTTAAAGAACTTCCGATTA 233
 OY 153 gaacacgctatggtattatgcaggtatataacttgaccacaatttggcgtagaacc 212
 DB 234 AACAATATGTTCTGGAACCTTGAAGGTTCAAGCCTGACAAAGATTAAGTAAATTAAC 293
 OY 213 cgaattgtgtgtcagaagcacaagaatttaatgca 249
 DB 294 AGTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 330

RESULT 11
 ID 004651 standard; DNA: 1915 BP.
 AC 004651;
 DR 02-OCT-1990 (first entry)
 DE OspA and OspB-encoding sequence of Borrelia burgdorferi
 DE Borrelia burgdorferi; Lyme disease; active immunisation;
 KW passive immunisation; OspA; OspB; ss.
 OS Borrelia burgdorferi.
 FH Key Location/Qualifiers
 FT rbs 1..138
 FT /tag= a
 FT /label= OspA RBS
 FT cds 151..972
 FT /tag= b
 FT /label= OspA structural gene
 FT -10_signal 102..107
 FT /tag= c
 FT /label= -10 region of P1
 FT -10_signal 108..113

FT /tag= d
 FT /label= -10 region of P2
 FT 80..85
 FT /tag= e
 FT /label= -35 region of P1
 FT 87..92
 FT /tag= f
 FT /label= -35 region of P2
 FT 972..977
 FT /tag= g
 FT /label= OspB RBS
 FT 982..1873
 FT /tag= h
 FT /label= OspB structural gene

FT WO9004411-A.
 FT 03-MAY-1990.
 PD 24-OCT-1989: W00248.
 PF 24-OCT-1988: DK-005902;
 PR (SYMB-) Symbicom AB.
 PA Bergstrom S, Barbour AG, Magnarelli LA;
 PI WPI: 90-163873/21.
 DR P-PSDB: R05028, R05029
 PT New immunologically active fractions of Borrelia burgdorferi -
 PT and derived antibodies and DNA coding sequences for active and
 PT passive immunisation against and diagnosis of Lyme disease.
 PS Disclosure: 112pp: English.
 CC OspA and OspB genes are part of the same operon located on a linear
 CC plasmid of B. burgdorferi and encode outer membrane proteins. They
 CC are separated by 12 bases within which is located the RBS of OspB.
 CC There are two putative promoters (P1 and P2) upstream of OspA. OspB
 CC include two closely spaced direct repeats of a 12-mer sequence
 CC (start-13 and 29) and a 14-mer palindromic sequence (start-123) which
 CC surrounds the P1 and P2 -10 regions.
 SQ Sequence 1915 BP; 811 A; 279 C; 337 G; 488 T;

Query Match 7.3%; Score 39.4; DB 1; Length 1915;
 Best Local Similarity 48.8%; Pred. No. 0.012;
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 33 agctcttcgtatgtgagatgagtgctaacgtccatcaactatggaattctgctga 92
 DB 1143 AGACTCAGTGTCTTGTATTAAAGTAATTAATTTTGTAAACCAAGAAAATAAGCTC 1202
 OY 93 tgcacacctatgtgtgccaatgtgcaagtacagtcgaagcgcaacaatcaacgtaa 152
 DB 1203 CGGCAAAATATGATTTAAAGCAACAATTCATCAGTTGAACCTTAAAGAACTTCCGATTA 1262
 OY 153 gaacacgctatggtattatgcaggtatataacttgaccacaatttggcgtagaacc 212
 DB 1263 AACAATATGTTCTGGAACCTTGAAGGTTCAAGCCTGACAAAGATTAAGTAAATTAAC 1322
 OY 213 cgaattgtgtgtcagaagcacaagaatttaatgca 249
 DB 1323 AGTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 1359

RESULT 12
 ID V00681 standard; DNA: 1916 BP.
 AC V00681;
 DR 08-JUN-1998 (first entry)
 DE Borrelia burgdorferi ospA and ospB gene region.
 DE Borrelia; antigen; outer surface protein; OspA; OspB; Lyme disease;
 KW DNA vaccine; immunisation; VR2210; VR2211; vector; ss.
 OS Borrelia burgdorferi strain B31.
 FH Key Location/Qualifiers
 FT RBS 139..144
 FT /tag= a
 FT /label= CDS
 FT 152..973
 FT /tag= b
 FT /label= sig_peptide
 FT 152..199
 FT /tag= c


```

FT mat_peptide 200..970
FT /*tag= d
FT 973..979
FT RBS /*tag= e
FT /983..1873
FT CDS /*tag= f
FT /983..1027
FT sig_peptide /*tag= g
FT /1028..1870
FT mat_peptide /*tag= h
FT
FT WO9747197-A1.
FT 18-DEC-1997.
FT 03-JUN-1997: U09439.
FT 14-JUN-1996: US-663998.
FT (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
FT (UTTE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
FT (VICA-) VICAL INC.
FT Barbour AG, Carner KR, Huebner RC, Liang X, Luke CJ,
FT Norman JA:
FT WPI: 98-051938/05.
FT
FT P-PSDB: W26771 AND W37131.
FT Plasmid for expressing Borrelia antigen in eukaryotic cells - used
FT as vaccines to protect against Lyme disease and for production of
FT antigens, themselves used in vaccines
FT
FT Example 1: Fig 7A-H: 80pp: English.
FT
FT This nucleotide sequence comprises the ospA and ospB gene region
FT of Borrelia burgdorferi. The ospA and ospB genes respectively
FT code for outer surface proteins (Osp) OspA (see W37131) and OspB
FT (see W26771) of B. burgdorferi. The ospA and ospB genes have
FT been used in the construction of VR2210 (see V00676) and VR2211
FT (see V00677). DNA plasmids suitable for transfection of e.g. baby
FT hamster kidney and human melanoma cells. Plasmid DNA encoding at
FT least one Borrelia genospecies antigen is disclosed and claimed.
FT The genospecies may be B. burgdorferi, Borrelia garinii and/or
FT Borrelia afzelii. The antigen can be OspA and/or OspB and/or
FT OspC. The plasmid is used in vaccines to elicit a protective immune
FT response (both humoral and cellular) in hosts, human or animal,
FT susceptible to Lyme disease (claimed). It can also be used for the
FT production of a Borrelia antigen in vitro in eukaryotic cells
FT (claimed), and these antigens used as vaccines or immunogenic
FT compositions, or to produce monoclonal antibodies.
FT
FT Sequence 1916 BP: 810 A: 280 C: 337 G: 489 T:
FT
Query Match 7.3%; Score 39.4; DB 1; Length 1916;
Best Local Similarity 48.8%; Pred. No. 0.012;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 33 agcttcctcgatattggcgatgagtgctaacgctgcacatcgctatggcaattcgtcga 92
DB 1144 AGACTCACTGCTCTTGTGTTAATGTAATAAAATTTTGTAGCAAAAGAAAATAGCTC 1203
QY 93 tgcataacctatgttggtgcacaattggtcaagtagacgcacaagcaatcaacggttaa 152
DB 1204 CGGCAAAATATGATTTAAGAGCAACATTCATCAGTTCACTTAAGAGAACTCCGATTA 1263
QY 153 gaacacccgtatggtattatgcaggtataacttgaccacaaatttggcgtgaacc 212
DB 1264 AAACAATGTTCTGGAACCCCTTGAAGGTTCAAAAGCCTGACAAAGATTAAGTAAATTAA 1323
QY 213 cgaatttggttcagacgcgcacaagaatttaagca 249
DB 1324 AGTTTCTGCTGATTTAAACACAGTAACCTTAGAACCA 1360

```

RESULT 13
V00677
ID V00677 standard; DNA: 5898 BP.
AC V00677;
DT 08-JUN-1998 (first entry)
DE DNA plasmid VR2211 containing Borrelia burgdorferi OspB gene.
KW Borrelia; antigen; outer surface protein A; OspB; Lyme disease;
KW DNA vaccine; immunisation; VR2211; vector; ss.

```

OS Chimeric - Borrelia burgdorferi strain B31.
OS Chimeric - Homo sapiens.
FT Key Location/Qualifiers
FT primer_bind complement (2..19)
FT /*tag= a
FT /note= "binding site of forward primer used to
FT amplify and insert OspB coding sequence"
FT primer_bind 835..852
FT /*tag= b
FT /note= "binding site of reverse primer used to
FT amplify and insert OspB coding sequence"
FT
FT WO9747197-A1.
FT 18-DEC-1997.
FT 03-JUN-1997: U09439.
FT 14-JUN-1996: US-663998.
FT (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
FT (UTTE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
FT (VICA-) VICAL INC.
FT Barbour AG, Carner KR, Huebner RC, Liang X, Luke CJ,
FT Norman JA:
FT WPI: 98-051938/05.
FT
FT Plasmid for expressing Borrelia antigen in eukaryotic cells - used
FT as vaccines to protect against Lyme disease and for production of
FT antigens, themselves used in vaccines
FT
FT Example 1: Fig 4A-D: 80pp: English.
FT
FT This is the nucleotide sequence of VR2211, a DNA plasmid suitable
FT for transfection of e.g. baby hamster kidney and human melanoma
FT cells. VR2211 contains DNA encoding the Borrelia burgdorferi
FT outer surface protein B (OspB). It was made by ligating: an OspB
FT fragment (see V00680) amplified from pTRH46 (see V00685-86); a
FT tissue plasminogen activator 5'-untranslated region/leader peptide
FT sequence amplified from nKCVMinIBL (see V00678); and pSLT- and
FT XbaI-digested plasmid VR1012. Plasmid DNA encoding at least one
FT Borrelia genospecies antigen is disclosed and claimed. The
FT genospecies may be B. burgdorferi, Borrelia garinii and/or
FT Borrelia afzelii. The antigen can be OspA and/or OspB and/or
FT OspC. The plasmid is used in vaccines to elicit a protective immune
FT response (both humoral and cellular) in hosts, human or animal,
FT susceptible to Lyme disease (claimed). It can also be used for the
FT production of a Borrelia antigen in vitro in eukaryotic cells
FT (claimed), and these antigens used as vaccines or immunogenic
FT compositions, or to produce monoclonal antibodies.
FT
FT Sequence 5898 BP: 1643 A: 1402 C: 1380 G: 1473 T:
FT
Query Match 7.3%; Score 39.4; DB 1; Length 5898;
Best Local Similarity 48.8%; Pred. No. 0.021;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 33 agcttcctcgatattggcgatgagtgctaacgctgcacatcgctatggcaattcgtcga 92
DB 118 AGACTCACTGCTCTTGTGTTAATGTAATAAAATTTTGTAGCAAAAGAAAATAGCTC 177
QY 93 tgcataacctatgttggtgcacaattggtcaagtagacgcgcacaagcaatcaacggttaa 152
DB 178 CGGCAAAATATGATTTAAGAGCAACATTCATCAGTTCACTTAAGAGAACTCCGATTA 237
QY 153 gaacacccgtatggtattatgcaggtataacttgaccacaaatttggcgtgaacc 212
DB 238 AAACAATGTTCTGGAACCCCTTGAAGGTTCAAAAGCCTGACAAAGATTAAGTAAATTAA 297
QY 213 cgaatttggttcagacgcgcacaagaatttaagca 249
DB 298 AGTTTCTGCTGATTTAAACACAGTAACCTTAGAACCA 334

```

RESULT 14
X20251/c
ID X20251 standard; DNA: 53585 BP.
AC X20251;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #4.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 11:05:58 ; Search time 278.23 Seconds
(without alignments)
247,802 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543
Sequence: 1 atgaatacttaataaacact.....gcgcctcattggtctttaa 543

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCBUS.COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.4	7.3	846	1 US-07-941-523-21	Sequence 21, Appl
2	39.4	7.3	891	1 US-07-941-523-19	Sequence 19, Appl
3	39.4	7.3	1959	2 US-08-137-175A-1	Sequence 1, Appl
4	39.4	7.3	5952	3 US-08-663-998-2	Sequence 2, Appl
5	31	5.7	2696	1 US-07-961-522-1	Sequence 1, Appl
6	31	5.7	2696	1 US-08-217-438-1	Sequence 1, Appl
7	31	5.7	2696	2 US-08-321-978-1	Sequence 1, Appl
8	31	5.7	2696	3 US-08-710-584-1	Sequence 1, Appl
9	31	5.7	246240	3 US-08-724-394A-20	Sequence 20, Appl
10	31	5.7	246240	3 US-08-724-394A-21	Sequence 21, Appl
11	31	5.7	246240	3 US-08-724-394A-22	Sequence 22, Appl
12	30.8	5.7	2520	4 US-08-450-351-1	Sequence 1, Appl
13	30.8	5.7	3762	5 US-08-772-270A-3	Sequence 3, Appl
14	30.8	5.7	4042	2 US-08-200-232-1	Sequence 1, Appl
15	30.8	5.7	4042	2 PCT-US95-02219-1	Sequence 1, Appl
16	30.8	5.7	4042	6 PCT-US95-02219-1	Sequence 1, Appl
17	30.8	5.7	4042	6 PCT-US95-02219A-1	Sequence 1, Appl
18	30.8	5.7	8370	4 US-08-488-706-1	Sequence 1, Appl
19	30.6	5.6	2278	1 US-08-258-188-1	Sequence 1, Appl
20	30.6	5.6	2278	2 US-08-526-813-1	Sequence 1, Appl
21	30.6	5.6	2278	6 PCT-US95-08554-1	Sequence 1, Appl
22	29.8	5.5	1903	1 US-07-961-522-3	Sequence 3, Appl
23	29.8	5.5	1903	1 US-08-217-438-3	Sequence 3, Appl
24	29.8	5.5	1903	2 US-08-321-978-3	Sequence 3, Appl
25	29.8	5.5	1903	3 US-08-710-584-3	Sequence 3, Appl
26	29.4	5.4	660	3 US-08-625-377-1	Sequence 1, Appl
27	29.2	5.4	1529	4 US-08-477-451-17	Sequence 17, Appl

C	28	29.2	5.4	1529	4 US-08-477-451-21	Sequence 21, Appl
	29	29.2	5.4	19932	4 US-08-477-451-25	Sequence 25, Appl
	30	28.8	5.3	2949	5 US-08-433-522A-1	Sequence 1, Appl
	31	28.8	5.3	3294	1 US-08-409-995-1	Sequence 1, Appl
	32	28.6	5.3	5319	2 US-08-169-927-1	Sequence 1, Appl
	33	28.4	5.2	2802	1 US-08-215-805A-79	Sequence 79, Appl
	34	28.4	5.2	3848	1 US-08-215-805A-1	Sequence 1, Appl
C	35	28.2	5.2	2329	4 US-08-982-232-13	Sequence 13, Appl
	36	28	5.2	211	2 US-08-435-684A-54	Sequence 54, Appl
C	37	28	5.2	211	4 US-08-934-877A-54	Sequence 54, Appl
	38	28	5.2	825	4 US-08-441-857-1	Sequence 1, Appl
	39	28	5.2	1020	4 US-08-441-857-5	Sequence 5, Appl
	40	28	5.2	1302	2 US-08-425-299A-1	Sequence 1, Appl
	41	28	5.2	1452	2 US-08-187-785-2	Sequence 2, Appl
C	42	27.8	5.1	1506	1 US-08-149-105-2	Sequence 2, Appl
C	43	27.8	5.1	1506	1 US-08-317-847-2	Sequence 2, Appl
	44	27.8	5.1	1988	2 US-08-469-202-26	Sequence 26, Appl
	45	27.8	5.1	1988	4 US-08-484-434C-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-07-941-523-21
Sequence 21, Application US/07941523
Patent No. 5571718
GENERAL INFORMATION:
APPLICANT: Dunn, John J
TITLE OF INVENTION: Cloning and Expression of Borrelia
TITLE OF INVENTION: Lipoproteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 01730
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/941,523
FILING DATE: 19920908
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BNL90-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-941-523-21

Query Match 7.3%; Score 39.4; DB 1; Length 846;
Best Local Similarity 48.8%; Pred. No. 0.0029;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 33 agctcttcgtatgtgagtgatgctcaagcgtccatcagctatgtgcaattcgtcgcga 92
Db 117 AGACTCAGTGTCTTGTGTTATGTAATTAATTTTGTATGCAAGAAAAAATATGCTC 176

[illegible]

```

RESULT 2
US-07-941-523-19
: Sequence 19, Application US/07941523
: Patent No. 5571718
: GENERAL INFORMATION:
: APPLICANT: Dunn, John J
: APPLICANT: Barbour, Alan G
: TITLE OF INVENTION: Cloning and Expression of Borrelia
: TITLE OF INVENTION: Lipoproteins
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millitia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 01730
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/941,523
: FILING DATE: 19920908
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: BNL90-01A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 861-6240
: TELEFAX: (617) 861-9540
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 891 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-07-941-523-19

```

Query Match	7.3%	Score	39.4	DB	1	Length	891
Best Local Similarity	48.8%	Pred. No.	0.0029				
Matches	106	Conservative	0	Mismatches	111	Indels	0
				Gaps			0

OY		33	agcttcttcgttaattggcga tgaagtgcatacgcgtgcacatcaagcatlcygcattctcgta	92
Db		162	AGACTCGTGTCTTTGTTTAATGGTATAAATAATTTTGTGAAGCAAAAGAAAAATATGC TC	221
OY		93	tgtccaacctatgttggtgccaaatttgtlcaagtagacyccaaacaatlcaacgtla	1522
Db		222	CGGCAATATGATTTAAGACCAACAATTGTATCAGTGATGAACTTAAAGCAACTCCCATPA	281
OY		153	gaacaccgcatatgynaattatigcagytataactttggacaanaatttggcgtagaac	212
Db		282	AACCATGTTCCTGGAACCTTTAAGGTTTAAAGCCTGACACAGTPAAGTAAAATTAAC	341

Qy 213 cgaattgtgtggtcagaagccaaagaatttaatgca 249
| | | | | | | | | | | | | | |
Db 342 agtttctgctgatttaaacacagtaaccttagaagca 378

```

RESULT      3
US-08-137-175A-1
: Sequence 1, Application US/08137175A
: Patent No. 577095
:
GENERAL INFORMATION:
APPLICANT: BARBOUR, Alan G.
APPLICANT: BERGSTROEM, Sven
APPLICANT: HANSSON, Lennart
TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDORFERI AND
TITLE OF INVENTION: PROPHYLAXIS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEWMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/137,175A
FILING DATE: 26-OCT-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCY/US92/08972
FILING DATE: 22-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, Iyer P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: BARBOUR-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
:
INFORMATION FOR SRO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1959 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: B31 (ATCC 35210)
:
FEATURE:
NAME/KEY: misc_feature
LOCATION: 123..142
OTHER INFORMATION: /function= "Primer"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 584..607
OTHER INFORMATION: /function= "Primer"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 776..794
OTHER INFORMATION: /function= "Primer"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 806..817
OTHER INFORMATION: /function= "Primer"
FEATURE:
NAME/KEY: CDS
LOCATION: 119..940
OTHER INFORMATION: /product= "ospA"

```

Query Match 5.7%; Score 31; DB 1; Length 2696;
Best Local Similarity 47.0%; Pred. No. 2.1;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1

QY	235	aaagaatttaatacagcgysaactccgtgaagaagysatgataacctttggtctat	294
Db	1662	ACAGAAATTTGATGTGATTTTGGCACAATAAATTTAAGTGGTAAGTTATTGATTA	1721
QY	295	ggccacatatgcataactatcatcaatbaccatttatgcaagggccaattagcaat	354
Db	1722	GGGTAAATCTGTGTTTACCGTGAAGTGGAAAAATTTGAAGTGAATGGTTTACTGGCAAA	1781
QY	355	gctaagactaaagtatggtttaccagccgtaatgcaactacatactcaacaagaagcac	414
Db	1782	GCTAAACCTCAATGAAGGCTTCGGCTCAAGTTCAAGTTACAGTTCACGTTATAGAAATGTG	1841
QY	415	aaaacacgctagcaagcggt	474
Db	1842	AAATTTAACATTTACACAGTTAAGTGGTGGCTTTCTAATGTCCACACGGCA---GCACAGCTT	1898
QY	475	gaagcaagctacaacatcatcatcaagaagaatgccaatgca	513
Db	1899	GCGCGACAAATTCACCATTAATTCAGAAATTTGGCAGTCGTA	1937

RESULT 6
US-08-217-438-1
; Sequence 1, Application US/08217438
; Patent No. 5521072

UNPUBLISHED INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Gerlach, Gerald F.
APPLICANT: Willson, Philip J.
APPLICANT: Rossi-Campos, Amalia
TITLE OF INVENTION: ACTINOBACILLUS PLEUROPEIMONTAE
TITLE OF INVENTION: TRANSFERIN BINDING PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Reed & Robins
STREET: 635 Bryant Street
City: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,438
FILING DATE: 22-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-617-8999
TELEFAX: 415-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
IS-08-217-438-1

Query Match	5.7%;	Score 31;	DB 1;	Length 2696;
Best Local Similarity	47.0%;	Pred. NO. 2.1;		
Matches 131;	Conservative	0;	Mismatches 145;	Indels 3;
				Gaps 1;

OY	235	aaagaatttaacgcagcgagctcctgttaaaagcgcgaagctctttggcgcttat	294
Db	1662	ACGAAATTATATGTGATTTTGGCAACAAAATTTTAAGTGGTAAGTATTGTGATTA	1721
OY	295	ggcacatcgcgtataactcatalcaatlacccattttagcacaaggcaaatltagcat	354
Db	1722	GGTGTAATCTGTTGTTTACCGTACGATGCAAAAATTTGATGTAAAGTGTCTAC	1781
OY	355	gctaaagactaaagtaagtgttaccagcgtaatgcacataactcaaaacaagcgaac	414
Db	1782	GCTAAACCTCAAGTGAAGGCTTCGCTTACGATTCAGGTAGTTCACGTTATGAGATG	1841
OY	415	aaaaccgcctiagcagcggt	474
Db	1842	AAATTTAACATGTACACAGTTAGTGGTGGCTCTTATAGTCCAAACGCA---GCA	1901
OY	475	gaagcaagctaacactatcatcataagaagttgcacaatgcga	513
Db	1899	GCGGACAAATTCACCACTAAATAGAAAATGGCAGGTA	1937

RESULT 7
US-08-321-978-1
; Sequence 1, Application US/08321978
; Patent No. 5801018
GENERAL INFORMATION

```

GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: APPLICANT: GERALD F.
APPLICANT: WILSON, PHILIP J.
APPLICANT: ROSSI-CAMPOS, AMALIA
TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
TITLE OF INVENTION: PLEUROPNEUMONIAE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ROBERTA L. ROBINS
STREET: 285 HAMILTON AVE, SUITE 200
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/321,978
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
OS-08-321-978-1

```

Query Match 5.7%; Score 31; DB 2; Length 2696;
Best Local Similarity 47.08; Pred. No. 2.1;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

[illegible]

```

1      RESULT 8
2      US-08-710-584-1
3      : Sequence 1, Application US/08710584
4      : Patent No. 5876725
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: POTTER, ANDREW A.
8      : APPLICANT: GERLACH, GERALD F.
9      : APPLICANT: WILSON, PHILIP J.
10     : APPLICANT: ROSSI-CAMPOS, ANA LIA
11     : TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
12     : TITLE OF INVENTION: PLEUROPEUMONIAE
13     : NUMBER OF SEQUENCES: 9
14     : CORRESPONDENCE ADDRESS:
15     : ADDRESSEE: ROBERTA L. ROBINSON
16     : STREET: 285 HAMILTON AVE, SUITE 200
17     : CITY: PALO ALTO
18     : STATE: CALIFORNIA
19     : COUNTRY: UNITED STATES OF AMERICA
20     : ZIP: 94301
21     :
22     : COMPUTER READABLE FORM:
23     : MEDIUM TYPE: Floppy disk
24     : COMPUTER: IBM PC compatible
25     : OPERATING SYSTEM: PC-DOS/MS-DOS
26     : SOFTWARE: PatentIn Release #1.0, Version #1.25
27     : CURRENT APPLICATION DATA:
28     : APPLICATION NUMBER: US/08/710,584
29     : FILING DATE:
30     : CLASSIFICATION:
31     : PRIOR APPLICATION DATA:
32     : APPLICATION NUMBER: US 08/321,978
33     : FILING DATE: 12-OCT-1994
34     : ATTORNEY/AGENT INFORMATION:
35     : NAME: ROBINSON, ROBERTA L.
36     : REGISTRATION NUMBER: 33,208
37     : REFERENCE/DOCKET NUMBER: 9001-0015.02
38     : TELECOMMUNICATION INFORMATION:
39     : TELEPHONE: (415) 327-3400
40     : TELEFAX: (415) 327-3231
41     : INFORMATION FOR SEQ ID NO: 1:
42     : SEQUENCE CHARACTERISTICS:
43     : LENGTH: 2696 base pairs
44     : TYPE: nucleic acid
45     : STRANDEDNESS: double
46     : TOPOLOGY: linear
47     : MOLECULE TYPE: DNA (genomic)
48     : FEATURE:
49     : NAME/KEY: CDS
50     : LOCATION: 333..1973
51     :
52     : US-08-710-584-1

```

[illegible]

```

1 RESULT 9
2 US-08-724-394A-20
3 : Sequence 20 Application US/08724394A
4 : Patent No. 5872237
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: Feder, John N.
8 : APPLICANT: Krommal, Gregory S.
9 : APPLICANT: Lauer, Peter M.
10 : APPLICANT: Ruddy, David A.
11 : APPLICANT: Thomas, Winston
12 : APPLICANT: Tsuchinashi, Zenta
13 : APPLICANT: Wolff, Roger K.
14 : TITLE OF INVENTION: Megabase Transcript Map: No. 5872237rel
15 : NUMBER OF SEQUENCES: 31
16 :
17 : CORRESPONDENCE ADDRESSES:
18 : ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
19 : STREET: Two Embarcadero Center, 8th Floor
20 : CITY: San Francisco
21 : STATE: CA
22 :
23 : COUNTRY: USA
24 : ZIP: 94111-3834
25 :
26 : COMPUTER READABLE FORM:
27 : MEDIUM TYPE: Floppy disk
28 : COMPUTER: IBM PC compatible
29 : OPERATING SYSTEM: PC-DOS/MS-DOS
30 : SOFTWARE: PatentIn Release #1.0, Version #1.30
31 :
32 : CURRENT APPLICATION DATA:
33 : APPLICATION NUMBER: US/08/724,394A
34 : FILING DATE: 01-OCT-1996
35 : CLASSIFICATION: 536
36 :
37 : ATTORNEY/AGENT INFORMATION:
38 : NAME: Pitts, Renee A.
39 : REGISTRATION NUMBER: 35,136
40 : REFERENCE/DOCKET NUMBER: 017957-000100
41 : TELECOMMUNICATION INFORMATION:
42 : TELEPHONE: 415-576-0200
43 : TELEFAX: 415-576-0300
44 : INFORMATION FOR SEQ ID NO: 20:
45 : SEQUENCE CHARACTERISTICS:
46 : LENGTH: 246240 base pairs
47 : TYPE: nucleic acid
48 : STRANDEDNESS: not relevant
49 : TOPOLOGY: not relevant
50 : MOLECULE TYPE: cDNA
51 : FEATURE:
52 : NAME/KEY: misc-feature
53 : LOCATION: 1..246240

```


Qy	16	acacattcggcagcatcagctcttcctcgatattggcgatagatgctcgaagctgcacacgc	75
Db	75473	ACACACTTGGCAACTCAACAACTTCAGATTTATTTCCGTGGCTGACCTAACCTCCACATTTATC	755322
Qy	76	tatggcaattcgcgcgatgcctcaccctatgtgtgtgcgaatatggtccagttaagcgc	135
Db	75533	AGAGCCACATCTCTTTATATGAAATATTATTAGGTGTTGTCAGAAAAGTAATATCGGCTCTTTGCC	755922
Qy	136	aagcaatcaacagcttaagaaacgc	158
Db	75593	ATTAAAGTAAAGCAGAAAAACAC	75615

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	5.7%	Score 30.8	DB 4	Length 2520
Best Local Similarity	42.8%	Pred. No. 2.3		
Matches 155	Conservative 0	Mismatches 207	Indels 0	Gaps 0
QY	91 gatgtcaaccctatggtgtgcacaatttgttcaagttagaagccaaagcaatacaagt	150		
DB	1408 GAAGCTGTAACCCCTAGAGGTTTAGACCTGGTTTTCACAAACCAAGAAAAACCATGCT	1467		
QY	151 aagaacacccgctatggtattttagcaggttaataacttgcacaaattttgctagaa	210		
DB	1468 GGCATTATCTAGAAAGCGGTTAACCAATCAATGACTTTACCATCAATACAGCGCGCT	1527		
QY	211 cccgaatttgtgtgtcagaagccaaagaatttaatgcaagcgtyagtcytaaaggt	270		
DB	1528 GTTGAACCGTTTGTGACTTTAAAGCTATGACGGCTTAAAGTTGGAAAAACCACTCAAC	1587		

QY	271	gatgtgaagctcttttggctgtcttatgycacatatcgctataacttcatcaatccccat	330
Db	1588	CCAAAGCTTTGGGGGATTTATGATGTCATATCCCTAATCTTAGCGTCACGGTAACCTAATC	1647
QY	331	tatgcgaagggcacaatttaggcattgtcctaagaactaaagtatgatgctacagccgtatbca	390
Db	1648	TATGCCACTGCGACCCACGCTTTGCTGATGCTATCTTAAGCCGTGGCTTCCGTATATGC	1707
QY	391	actacatactcaacaaacaaagcgacaaaacacagcttaccagcggtgtgtgtgtgtgtgt	450
Db	1708	GTTTGAGATATTGCTGATTAACGAAAGACGAAAAAGCAGCAATACGAGATTGTGTTT	1767
QY	451	aa 452	
Db	1768	AA 1769	

```

US-08-450-351-3
RESULT 13
Sequence 3, Application US/08450351
Patent No. 5981213
GENERAL INFORMATION:
APPLICANT: Hansen, Eric J.
APPLICANT: Helminen, Merja E.
APPLICANT: Maciver, Isobel
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO THE USEFUL ANTIGENS OF MORAXELLA
TITLE OF INVENTION: USEFUL ANTIGENS OF MORAXELLA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,351
FILING DATE: 25-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: AMCR:019
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2520 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 100..2376
US-08-450-351-3
```

Query Match	5.7%	Score 30.8	DB 4	Length 2520
Best Local Similarity	42.8%	Pred. No. 2.3		
Matches 155	Conservative	0	Mismatches 207	Indels 0
				Gaps 0
Qy	91	gatgtcaaccttaattgtgtgccaatattgttcaagtacgcgaagaacatacaagt	150	
Db	1408	GAAGCTGTACCCCGCCAGAGGATATTGACTGCTTTTCAAAACCAAGAAAAAACCAGTGT	1467	
Qy	151	aagaacaccgcttaattgtattattatgcagttataacttgcacaaattttgctgttagaa	210	

.Tue.May 23 16:14:44 2000

us-09-164-714-6.rni

Page 9

Db 2891 AATGAAATCACCAGCAATGTGATACGGCCACTACCACTTA 2932

Search completed: May 20, 2000, 11:09:40
Job time: 13402 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 10:35:43 ; Search time 5377.06 Seconds
(without alignments)
409.313 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543
Sequence: 1 atgaactttaaacact.....gcgcctattgcttttaa 543

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

```
EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	39.4	7.3	1101	83	CNS0182P	AL108811 Drosophila
2	38.6	7.1	344	80	AM334656	AM334656 S38H1 AGS
3	38.6	7.1	1101	82	CNS000EAX	AL068861 Drosophila
4	36.6	6.7	587	103	AQ256569	AQ256569 nbdx0016B
5	36.4	6.7	479	90	AQ826019	AQ826019 HS_5442_A
6	36.4	6.7	1101	82	CNS000F8	AL052505 Drosophila
7	35.8	6.6	460	42	AL138233	AL138233 qd7603.x
8	35.8	6.6	514	90	AQ784640	AQ784640 HS_3193.B
9	35.2	6.5	1000	82	CNS000A7D	AL054816 Drosophila
10	35	6.5	1000	82	CNS000CQ	AL059446 Drosophila
11	34.8	6.4	407	74	AM201163	AM201163 seq8f03.y
12	34.8	6.4	556	103	AQ176788	AQ176788 HS_3213_A
13	34.8	6.4	1086	82	CNS000YXK	AL069662 Drosophila
14	34.6	6.4	593	20	Z46831	Z46831 ATTS4393.Ra
15	34.6	6.4	598	60	AT194544	AT194544 fcd45f09.y
16	34.6	6.4	801	107	AQ449605	AQ449605 500002D09
17	34.4	6.3	402	102	AQ124099	AQ124099 HS_3122_A
18	34.4	6.3	442	74	AM202121	AM202121 s12b08.y
19	34.4	6.3	460	62	AL1900885	AL1900885 sb95f12.y
20	34.4	6.3	464	81	AM397695	AM397695 sg83b10.y
21	34.2	6.3	530	103	AQ214218	AQ214218 HS_3011_A
22	34.2	6.3	797	82	CNS003F8	AL064634 Drosophila
23	34	6.3	347	21	R02601	R02601 yef6a01.r1
24	33.8	6.2	374	32	AA375686	AA375686 EST88188
25	33.8	6.2	392	40	AA908681	AA908681 c104e04.s
26	33.8	6.2	565	35	AA557106	AA557106 948 l0blo
27	33.8	6.2	569	106	AQ384040	AQ384040 R0c11-15
28	33.6	6.2	380	80	AV403461	AV403461 AV403461
29	33.6	6.2	395	47	AL153238	AL153238 SD03803.5
30	33.4	6.2	568	85	AQ698184	AQ698184 HS_5547_B
31	33.4	6.2	1039	83	CNS017AD	AL107791 Drosophila
32	33.2	6.1	365	47	AL1497040	AL1497040 fb59d08.y
33	33.2	6.1	493	85	AQ718562	AQ718562 HS_5509_B
34	33.2	6.1	521	61	AL1827597	AL1827597 w10910.x
35	33.2	6.1	649	94	AQ003789	AQ003789 CpG0394B
36	33	6.1	300	35	C37766	C37766 C37766.YuJ1
37	33	6.1	508	35	AA553216	AA553216 VK8B09.s
38	33	6.1	510	33	AA408349	AA408349 EST02828
39	33	6.1	532	81	AM398432	AM398432 EST298279
40	33	6.1	1101	82	CNS012RM	AL101332 Drosophila
41	32.8	6.0	381	38	AA748003	AA748003 nx87b01.s
42	32.8	6.0	485	107	AQ476428	AQ476428 CTRBI-E1-
43	32.8	6.0	599	63	AL197458	AL197458 701554019
44	32.8	6.0	689	61	AT1897404	AT1897404 w198902.x
45	32.6	6.0	431	25	N97920	N97920 1539c3 czap

ALIGNMENTS

RESULT 1						
LOCUS	CNS0182P	1101 bp	DNA	GSS	26-JUL-1999	
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BAC37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
ACCESSION	AL108811					
VERSION	AL108811.1	GI:5629115				
KEYWORDS	GSS.					
SOURCE	fruit fly.					
ORGANISM	Drosophila melanogaster					
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
AUTHORS	1 (bases 1 to 1101)					
TITLE	Genoscope.					
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)					
COMMENT	Determination of this BAC-end sequence was carried out as part of a					

collaboration with the European Drosophila genome Project (EDGP) - <http://www.edgp.ebi.ac.uk>. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

source

1. 1101
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_11b="DrosBAC"
/clone="BAC37D10"
/note="end : SP6"
BASE COUNT 274 a 268 c 128 g 73 t 358 others
ORIGIN

Query Match

Best local similarity 18.6%; Pred. No. 0.36;
Matches 66; Conservative 138; Mismatches 151; Indels 0; Gaps 0;

QY	189	tgaccacaaatttgccgctgaacccgaatttggtctgaagcccaagaatttaagc	248
DB	662	DGCGAAADAWMRGRRARRAARAAADAKRAAAAAAAAAAAAAAAAAAGGRRK	721
QY	249	aggcgagatctcgtgaagaagtgatgtaagcttcttgctgaagcacatatgcc	308
DB	722	WKGDKDKGKATKTAAMKRGKRDGTATAMWTMDTAMADTAKRAAARRRR	781
QY	309	taactcacaatacccaattatgccaaagcaaatgaagcatgctgaagtaagtc	368
DB	782	DARKTRDGRARRRARRAAGGRARRARARRAARADRDADAAAAAAAAA	841
QY	369	agatgtaccagccgtaatgcaatacactaacaaagcgaacacagcctagc	428
DB	842	AAWTTWRDRMWDMDWDTTAAWMDARARARRRRRRRRARRAARADDT	901
QY	429	aggcgatgt	488
DB	902	DTKDRADATTTDDTTKTTTDDDDKAKRDMWMAKADGAWKMDRARDWAATK	961
QY	489	ctatcacaagaagatgcaatgattgtggcgctcatgtgctttta	543
DB	962	DGKMDGGRGRRGKDKKRMKDKKDDDDKXTWTWRDMWMTTKKMDMW	1016

RESULT 2						
LOCUS	AM334656/c	344 bp	mRNA	EST	31-JAN-2000	
DEFINITION	S38H1 AGS-1 Pneumocystis carinii f. sp. carinii cDNA 3', mRNA sequence.					
ACCESSION	AM334656					
VERSION	AM334656.1	GI:6831117				
KEYWORDS	EST.					
SOURCE	Pneumocystis carinii f. sp. carinii.					
ORGANISM	Pneumocystis carinii f. sp. carinii.					
REFERENCE	Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae; Pneumocystis.					
AUTHORS	1 (bases 1 to 344)					
TITLE	Smulan,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman,J.C., Kovacs,J. and Cushion,M.					
JOURNAL	Expressed sequence tags from Pneumocystis carinii					
COMMENT	Unpublished (2000) On Jun 22, 1998 this sequence version replaced gi:3246719. Contact: Staben C School of Biological Sciences University of Kentucky 101 Morgan Building, University of Kentucky, Lexington, KY 40506-0225, USA Tel: 606 257 2161 Fax: 606 257 1717					


```

/clone="hxb0016B19r"
/clone_lib="CGI Rice BAC Library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBelobAC11; Site 1: HindIII; Site 2:
HindIII. Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 kb providing 10.9 haploid genome equivalents.
The deep coverage allows the isolation a particular
sequence with a probability of 99.9%. Two high density
filters, each containing 18,432 clones (doubly spotted),
represent the whole library for colony screening."

BASE COUNT      183 a      128 c      112 g      164 t
ORIGIN

Query Match
Best Local Similarity 56.1%; Pred. No. 2;
Matches 69; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 309 taactcatcaatacccatcttattgccaaggaataaggcattgctaagactaaagt 368
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 460 TCAATTAATGTTGATTCGCAATGATGCCAGACACTAATGATGATTCCTAGACACAAAG 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 369 agatgtaccacgcggaatgaagcaactatcctcaacaaagcagaacacagcctagc 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 520 AANCGCAATCAATCGCAATCGACCCACATGACAGCAAAACCCCTGATCAGCTAGC 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 429 agg 431
    |||
DB 580 ACG 582

RESULT 5
A0826019 479 bp DNA GSS 27-JUN-1999
LOCUS HS-5442_A1-G03-SP6E-RPCT-11 Human Male BAC Library Homo saplans
DEFINITION genomic clone Plate=1018 Col=5 Row=M, genomic survey sequence.
ACCESSION A0826019
VERSION A0826019.1 GI:5792081
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 479)
Mahtiras,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahtiras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCT-11. For BAC
library availability, please contact Pieter de Jong

```

```

(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1018 Row: M Column: 5
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 479.
Location/Qualifiers
1..479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1018 Col=5 Row=M"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACe3.6; Genomic sequence of BAC ends"

BASE COUNT      151 a      83 c      98 g      145 t      2 others
ORIGIN

Query Match
Best Local Similarity 53.5%; Pred. No. 2;
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 82 aattctgtcatgtcgaaccctatgttgcccaaatggtcgaagtagacgccaaagca 141
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 335 ATTTCGCTTGAGAGCTCTCTGTTATTTTGTATCAAAATGGAAGCAAAAATTCACAGCAG 394
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 142 atcaagggtaagacacgcctatgtattatgcaggtatataacttaccaaatttt 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 395 ATGACACAAAGAAAGGGGCAAAAGCTCTATGACTGTAAACAGACATCTTAGTAT 454
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 202 ggcgtagaaccgcgaattgtgt 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 455 GGTATCTAATTAAGGTCTATGCG 476

RESULT 6
CNS008F6/c
LOCUS CNS008F6/c
DEFINITION
Drosophila melanogaster genome survey sequence T713 end of BAC #
BACR17N15 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL052505
VERSION AL052505.1 GI:4933656
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers

```

FEATURES

```
source
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_id="RPC1-98"
/clone="BACR17N15"
/Note="end : TET3"
BASE COUNT      305 a      153 c      198 g      321 t      124 others
ORIGIN

Query Match
Best Local Similarity 39.6%; Pred. No. 2.8;
Matches 80; Conservative 25; Mismatches 97; Indels 0; Gaps 0;

OY 264 aaaaagtgaatgaagcttctgtgtatgacacatcgtatcaatcaatcaac 323
Db 923 AAAATTTATGKGMMATTTATGTTTTATRAAARAAAATAATMCCCATTTTCT 864
OY 324 cccattatgcgaaggcaattagcattgctgaagcctaagtagtattacagccg 383
Db 863 GSCATWTTCAATCCCTCYCAAMAANKGATTTATATCCCTAAATAATGTTGCAAAAT 804
OY 384 taatgcacatactacatacaaaagcgacaaacacagcctagcagcggtgtgtgt 443
Db 803 TCTWAGTHTCAMATTAAMAMATGMAAAMAAAMATCTGTTGTTWAKGGGTTTCTAT 744
OY 444 tgccttaaccattagcaat 465
Db 743 TTKGGTGACACATACYSCAAY 722

RESULT 7
A1138233/c 460 bp mRNA EST 28-OCT-1998
LOCUS      gd76e03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1735420
DEFINITION 3', mRNA sequence.
ACCESSION  A1138233
VERSION    A1138233.1 GI:3644205
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 460)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On May 1, 1998 this sequence version replaced gi:2339570.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            CDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILMIL at:
            www-bio.llnl.gov/bdrrp/image/image.html

Insert Length: 1159 Std Error: 0.00
Seq Primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 451.
Location/Qualifiers
1. .460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1735420"
/clone_id="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/Note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

```
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTTCACCAATCTGAAGTGGAGCGCGCCGCCCAATTTTTTTTTTTT 3'}.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      127 a      94 c      96 g      143 t
ORIGIN

Query Match
Best Local Similarity 54.1%; Pred. No. 3.3;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 403 aacaaagcgacaaacacagcctagcagcggtgtgtgtgtgttaaacattagca 462
Db 304 AAGAAGCTCCCAAAAACAGTCTGGAAGTCAGAGAGATATGCTCAACTACTTGAT 245
OY 463 aatgtggcggttgaagcaagctacactatctatcagaagatgccaatgcaattgtg 522
Db 244 GGCTTTACCGAATTTTTCAGAGACATCTATGTCATATAAAGCAATGCACTGACTAT 185
OY 523 ggcgtcaattgtgct 537
Db 184 CGCTCTCAAGTGCTT 170

RESULT 8
AO784640/c 514 bp DNA GSS 03-NOV-1999
LOCUS      HS_3193_B1 B07 T7C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-3193 Col-13 Row-D, genomic survey
ACCESSION  AO784640
VERSION    AO784640.1 GI:5692264
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 514)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.hsc.washington.edu
            Plate: 3193 Row: D Column: 13
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 514.
            Location/Qualifiers
1. .514
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3193 Col-13 Row-D"
/clone_id="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/Note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
```


Query Match	Best Local Similarity	Matches	Score	DB	Length
6.4%	54.8%	69	34.8	74	407
Conservative	0	Mismatches	57	Indels	Gaps
0	0	0	0	0	0
171	ttatgcaggtataacttgaccataaatttgctgctagaaaccggaattgtgtgcaga	230			
224	TAATACACTGTTAATTTTGAACGCTATCTTGTGGAGAAATCCTAACACTGTTGGTGCTGA	283			
231	cgcacaagaatttaatgcagcgctgtagtcctgtlaaagagtgtatgtaagtcctttgtgac	290			
284	GGCTGGAACAACAAAGCGTGGACAAAGGTAATCTTCACTGCAATGCAATGCTTAATGAGT	343			
291	ttatgg	296			
344	GGATTG	349			
BASE COUNT	116 a	84 c	94 g	113 t	
ORIGIN					
Query Match	Best Local Similarity	Matches	Score	DB	Length
6.4%	54.8%	69	34.8	74	407
Conservative	0	Mismatches	57	Indels	Gaps
0	0	0	0	0	0
171	ttatgcaggtataacttgaccataaatttgctgctagaaaccggaattgtgtgcaga	230			
224	TAATACACTGTTAATTTTGAACGCTATCTTGTGGAGAAATCCTAACACTGTTGGTGCTGA	283			
231	cgcacaagaatttaatgcagcgctgtagtcctgtlaaagagtgtatgtaagtcctttgtgac	290			
284	GGCTGGAACAACAAAGCGTGGACAAAGGTAATCTTCACTGCAATGCAATGCTTAATGAGT	343			
291	ttatgg	296			
344	GGATTG	349			
BASE COUNT	116 a	84 c	94 g	113 t	
ORIGIN					
Query Match	Best Local Similarity	Matches	Score	DB	Length
6.4%	54.8%	69	34.8	74	407
Conservative	0	Mismatches	57	Indels	Gaps
0	0	0	0	0	0
171	ttatgcaggtataacttgaccataaatttgctgctagaaaccggaattgtgtgcaga	230			
224	TAATACACTGTTAATTTTGAACGCTATCTTGTGGAGAAATCCTAACACTGTTGGTGCTGA	283			
231	cgcacaagaatttaatgcagcgctgtagtcctgtlaaagagtgtatgtaagtcctttgtgac	290			
284	GGCTGGAACAACAAAGCGTGGACAAAGGTAATCTTCACTGCAATGCAATGCTTAATGAGT	343			
291	ttatgg	296			
344	GGATTG	349			
BASE COUNT	116 a	84 c	94 g	113 t	
ORIGIN					
Query Match	Best Local Similarity	Matches	Score	DB	Length
6.4%	54.8%	69	34.8	74	407
Conservative	0	Mismatches	57	Indels	Gaps
0	0	0	0	0	0
171	ttatgcaggtataacttgaccataaatttgctgctagaaaccggaattgtgtgcaga	230			
224	TAATACACTGTTAATTTTGAACGCTATCTTGTGGAGAAATCCTAACACTGTTGGTGCTGA	283			
231	cgcacaagaatttaatgcagcgctgtagtcctgtlaaagagtgtatgtaagtcctttgtgac	290			
284	GGCTGGAACAACAAAGCGTGGACAAAGGTAATCTTCACTGCAATGCAATGCTTAATGAGT	343			
291	ttatgg	296			
344	GGATTG	349			
BASE COUNT	116 a	84 c	94 g	113 t	
ORIGIN					
Query Match	Best Local Similarity	Matches	Score	DB	Length
6.4%	54.8%	69	34.8	74	407
Conservative	0	Mismatches	57	Indels	Gaps
0	0	0	0	0	0
171	ttatgcaggtataacttgaccataaatttgctgctagaaaccggaattgtgtgcaga	230			
224	TAATACACTGTTAATTTTGAACGCTATCTTGTGGAGAAATCCTAACACTGTTGGTGCTGA	283			
231	cgcacaagaatttaatgcagcgctgtagtcctgtlaaagagtgtatgtaagtcctttgtgac	290			
284	GGCTGGAACAACAAAGCGTGGACAAAGGTAATCTTCACTGCAATGCAATGCTTAATGAGT	343			
291	ttatgg	296			
344	GGATTG	349			
BASE COUNT	116 a	84 c	94 g	113 t	
ORIGIN					
Query Match	Best Local Similarity	Matches	Score	DB	Length
6.4%	54.8%	69	34.8	74	407
Conservative	0	Mismatches	57	Indels	Gaps
0	0	0	0	0	0
171	ttatgcaggtataacttgaccataaatttgctgctagaaaccggaattgtgtgcaga	230			
224	TAATACACTGTTAATTTTGAACGCTATCTTGTGGAGAAATCCTAACACTGTTGGTGCTGA	283			
231	cgcacaagaatttaatgcagcgctgtagtcctgtlaaagagtgtatgtaagtcctttgtgac	290			
284	GGCTGGAACAACAAAGCGTGGACAAAGGTAATCTTCACTGCAATGCAATGCTTAATGAGT	343			
291	ttatgg	296			

Db	324	TGCTGATGAGGNNAGATGAGGAGGAAGATGAGAAAGACATGCTTTGAAACGAGAAGCTT	325
Oy	165	Tgtaattatbvcaggtactataacatttgacccaaaatttggcgttagaacccgaattbgttg	224
Db	264	TGAAGATGACGAAGAAGACATGCTTTGAGCACTA-TTTGGCGGAGAAATCGATCTTCGC	266
Oy	225	ttcagaagccaaagaatttaattgcaagcgtgagtcctcgtaaaagvtgactgaagtcctt	284
Db	205	CTCCCCCAACAAGACATGAAGAGCTTAAGGAGAGGCTTATATACGATAAATTTGAAGCTTT	146
Oy	285	tgatgctatgycacatatacgtcttaacttcac	318
Db	145	TTCTGGATGAGATCATGATGCAGATACACTGACC	112

RESULT	15
LOCUS	A1794544
DEFINITION	A1794544 598 bp mRNA EST 02-JUL-1999 fc5f109.1 zebrafish washu MIMC EST Danio rerio cDNA 5', mRNA sequence.
ACCESSION	A1794544
VERSION	A1794544.1 GI:5342260
KEYWORDS	EST.
SOURCE	zebrafish.
ORGANISM	Danio rerio

REFERENCE
AUTHORS

Clark, M., Johnson, S. L., Lehrnach, H., Lee, R., Li, F., Marra, M.,
1 (bases 1 to 598)
Cyrillinoidea, Cyrillinae, Rasporinae; Dario.
Euparietia, Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Osteariophysi; Cyrilliformes;
Cyrillinoidea, Cyrillinae, Rasporinae; Dario.

TITLE
 JOURNAL
 COMMENT

Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Smaller, T., Gibbons, M., Page, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Washu zebrafish EST project 1998
 On Jun 5, 1998 this sequence version replaced g1:3189700.

Other_Estrs: fc45f09.xl
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbraflshewatson.wustl.edu
cDNA Library Preparation by: Matthew Clark, cDNA Library Arrayed by:
Matthew Clark, DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourceCenter@umrprimardatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 498.
Location/Qualifiers
1..598

```

/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish Washu MPING EST"
/sex="mixed"
/tissue-type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLL-blue MRF"
/notes="vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
15'GGACTAGTTCAGATCGGAGCGCGCCGCTTTTCTTTTCTT3';
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrbach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysts were selected following oligonucleotide

```

hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

Query Match	6.4%	Score 34.6	DB 60	Length 598
Best Local Similarity	49.2%	Pred. No. 8		
Matches	91	Conservative	0	Mismatches 94
				Indels 0
				Gaps 0
QY	145	aacggtgaagAACaccgctatgtaattatcaggttaaacctttgacaaathtggc	204	
Db	181	AACATTTGAAATATGCGATACGAAATGATTTTGAGTTTAAACGCCAAAAAATATATC	240	
QY	205	gtagagaccggaattggtgtgtcagaagcaagaaattatgtcagcgggtgaagtcctta	264	
Db	241	AAAAAACTCAAGTATCTCTTTTAAACAAAAGATTTTCTCCCTTGCCTTTTACATCA	300	
QY	265	aaagtgatgtgaagctctttgtgtcattatgcacatacgcataaacctatcatcaatcc	324	
Db	301	AATGTACAAATTTTGTCATGTTAGTGGATGCCATTATATGGATAACTGCTTGCTATA	360	
QY	325	ccatt	329	
Db	361	CAGTT	365	

Search completed: May 20, 2000, 10:35:50
Job time: 12797 sec

‘ . ,

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 11:13:12 ; Search time 34.13 Seconds

(Without alignments)
124,919 Million cell updates/sec

Title: US-09-164-714-7

Sequence: 1 MKTKTLTLLAVASALLAMSA.....SYNTLSEDAANISLGAHLAF 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	12.6	188	1	R26415
2	112.5	12.3	361	1	R05999
3	112.5	12.3	361	1	R66877
4	112	12.3	359	1	R66294
5	112	12.3	362	1	P90098
6	108	11.8	188	1	R25674
7	108	11.8	188	1	R70744
8	108	11.8	188	1	W18380
9	106.5	11.7	162	1	R96207
10	106.5	11.7	338	1	R85450
11	106.5	11.7	568	1	R96210
12	100	11.0	288	1	W27720
13	100	11.0	1287	1	R79944
14	97	10.6	211	1	W20755
15	97	10.6	211	1	W55324
16	95.5	10.5	1290	1	W98269
17	95	10.4	1288	1	W55547
18	95	10.4	1288	1	W55685
19	94.5	10.4	330	1	W21678
20	93.5	10.3	257	1	W55462
21	91	10.0	1296	1	R41198
22	89.5	9.8	767	1	W46272
23	89.5	9.8	1267	1	R04232
24	88.5	9.7	174	1	W04891
25	87.5	9.6	443	1	W94460
26	86.5	9.5	363	1	R66878
27	85.5	9.4	711	1	W73025
28	85.5	9.4	758	1	W46270
29	85	9.3	621	1	W55545
30	85	9.3	637	1	W55739
31	85	9.3	691	1	W73027
32	85	9.3	691	1	W73032
33	84.5	9.3	255	1	W55201
34	84.5	9.3	255	1	W55477

35	84	9.2	174	1	W04893	Proteinase K resis
36	83	9.1	759	1	W46271	Moraxella catarrha
37	83	9.1	759	1	W46273	Moraxella catarrha
38	83	9.1	1992	1	W04505	Moraxella 200 kDa
39	82.5	9.0	595	1	W98283	H. pylori GHPO 125
40	80.5	8.8	341	1	R07042	P3 gene product of
41	80.5	8.8	342	1	R66879	H. influenzae prot
42	80.5	8.8	342	1	R66880	H. influenzae prot
43	80.5	8.8	1612	1	W65088	R. prowazekii S-la
44	79.5	8.7	175	1	W04892	Proteinase K resis
45	79.5	8.7	483	1	W73031	Helicobacter pylor

ALIGNMENTS

RESULT	ID	Score	Match	Length	DB ID	Description
1	R26415	115	12.6	188	1	R26415
AC	R26415	112.5	12.3	361	1	R05999
DE	R26415	112.5	12.3	361	1	R66877
KW	R26415	112	12.3	359	1	R66294
KM	R26415	112	12.3	362	1	P90098
OS	R26415	108	11.8	188	1	R25674
FH	R26415	108	11.8	188	1	R70744
FT	R26415	108	11.8	188	1	W18380
FT	R26415	106.5	11.7	162	1	R96207
FT	R26415	106.5	11.7	338	1	R85450
FT	R26415	106.5	11.7	568	1	R96210
FT	R26415	100	11.0	288	1	W27720
FT	R26415	100	11.0	1287	1	R79944
FT	R26415	97	10.6	211	1	W20755
FT	R26415	97	10.6	211	1	W55324
FT	R26415	95.5	10.5	1290	1	W98269
FT	R26415	95	10.4	1288	1	W55547
FT	R26415	95	10.4	1288	1	W55685
FT	R26415	94.5	10.4	330	1	W21678
FT	R26415	93.5	10.3	257	1	W55462
FT	R26415	91	10.0	1296	1	R41198
FT	R26415	89.5	9.8	767	1	W46272
FT	R26415	89.5	9.8	1267	1	R04232
FT	R26415	88.5	9.7	174	1	W04891
FT	R26415	87.5	9.6	443	1	W94460
FT	R26415	86.5	9.5	363	1	R66878
FT	R26415	85.5	9.4	711	1	W73025
FT	R26415	85.5	9.4	758	1	W46270
FT	R26415	85	9.3	621	1	W55545
FT	R26415	85	9.3	637	1	W55739
FT	R26415	85	9.3	691	1	W73027
FT	R26415	85	9.3	691	1	W73032
FT	R26415	84.5	9.3	255	1	W55201
FT	R26415	84.5	9.3	255	1	W55477

35	84	9.2	174	1	W04893	Proteinase K resis
36	83	9.1	759	1	W46271	Moraxella catarrha
37	83	9.1	759	1	W46273	Moraxella catarrha
38	83	9.1	1992	1	W04505	Moraxella 200 kDa
39	82.5	9.0	595	1	W98283	H. pylori GHPO 125
40	80.5	8.8	341	1	R07042	P3 gene product of
41	80.5	8.8	342	1	R66879	H. influenzae prot
42	80.5	8.8	342	1	R66880	H. influenzae prot
43	80.5	8.8	1612	1	W65088	R. prowazekii S-la
44	79.5	8.7	175	1	W04892	Proteinase K resis
45	79.5	8.7	483	1	W73031	Helicobacter pylor

Query Match 12.6%; Score 115; DB 1; Length 188;
Best Local Similarity 27.1%; Pred. No. 0.0005;
Matches 55; Conservative 28; Mismatches 82; Indels 38; Gaps 11;

QY	1	MKTTLTLLAVASALLAMSA	SYNSADAOPRYGAKTIOVDAKJNGKNTAYGIYAG	60
DB	1	MKNTILSLVLTSTLYLVNNA	QADOTMAESVYABYQSKV--QPFKNIRGVNRY-----R	54
QY	61	YNFQONFGEAFVGS	DA-----KEFNAGVSPV-----KGDVSEFGAYGT-----YRY	103
DB	55	YEDDS-----PYSFISLS	LYLVDRQASGSVPEGIHYDKREVK-----YGLMNGPARYRL	106
QY	104	NFINTPYAKGKLGIAK-TKVDVTSRNATYYSNR--SDKTSIAGGVGVGFPLANVGEYA		160
DB	107	S-DNFSLYALAGVGTVAATFEKHESTQDDSPSNKISRKTFANGAGVGMPLNIVVDV		165
QY	161	SY---NTLSEDAANISLGAHLAF		180

Query Match 12.3%; Score 112; DB 1; Length 359;
Best Local Similarity 28.0%; Pred. No. 0.0023;
Matches 55; Conservative 18; Mismatches 80; Indels 42; Gaps 11;

QY 5 KTLIAVSASSILMSANAISYGSNADAPYVGAKIGQ-----VDKQING----- 51
DB 3 KTLIAVYAGLAAASVQAAPQENTF-----YAGYKAGOGSFDHCINNNGAIKKGLSSSNY 58
QY 51 ---KNT-AYGIAGYNE--DQNGVEAEFVGSQ---AKFNAGVSPYKGDYKSPGAY--- 99
DB 59 GYRNFTFTYGVFGGYQLNDQNFCLAEELGHDGRRAKLEBAGKPKAKH--TNHGATLSL 116
QY 99 -GYRYNFINTPYFAKCKLGIAKTKVD---VTSRNATYTSKSKTSLAGG---VGYGFK 151
DB 117 KGSE---VLDGLDVGKAGVALRSDYKFEYEDANGTRDHRKGRHTARASGLFAVGAIEYA 173
QY 152 PLAVGVEASYNYLS 166
DB 174 VLPELAVRLEYQWLT 188

RESULT 5
P90098
ID P90098 standard; protein; 362 AA.
AC P90098;
DT 1-NOV-1989 (first entry)
DE P2 antigen of Haemophilus influenzae
KW Haemophilus influenzae P2 antigen; vaccines; pathogenic
KW type b strains.
OS Haemophilus influenzae type b
PN EP-320289-A.
PD 14-JUN-1989.
PF 09-DEC-1988; 311691.
PR 10-DEC-1987; US-249482.
PA (TEXA) Univ of Texas Syst.
PI Hansen ET.
DR WPI: 89-174562/24.
PT DNA encoding Haemophilus influenzae P2 antigen
PT - used for producing immunogenic comps. as vaccines
PT and in diagnostic hybridisation assays.
PS Disclosure; fig 6; 20pp; English.
CC P2 antigen of Haemophilus influenzae (see corresp. N90033).
CC Isolated from type b strains, the P2 antigens are used as vaccines
CC and to detect P2-specific RNA or DNA. Segments used as antigens
CC Include amino acids 21-361, 175-197, 260-275, 296-311, 33-353,
CC or their functional equivalents.
SQ Sequence 362 AA;

R25674
ID R25674 standard; protein; 188 AA.
AC R25674;
DT 13-JAN-1993 (first entry)
DE pagC.
KW Salmoneilla positive regulon; phoP; vaccination.
OS Salmoneilla typhimurium.
PN W09211361-A.
PD 09-JUL-1992.
PF 18-DEC-1991; U09604.
PR 18-DEC-1990; US-629602.
PA (GEHO) GEN HOSPITAL CORP.
PA (HARD) HARVARD COLLEGE.
PI Mekalanos JT, Miller SI;
DR WPI: 92-250084/30.
DR N-PSDB; Q26694.
PT New anti-salmoneilla vaccine contg. live attenuated mutants -
PT useful to protect against typhoid fever and related diseases, and
PT for detecting salmoneilla
PS Disclosure; Page 45-51; 57pp; English.
CC pagC is a Salmoneilla positive regulon (phoP) regulated gene. A
CC mutation at pagC confers a virulence defect. Within the scope of
CC the invention, this gene was mutated, such that it included a
CC virulence attenuating mutation. This mutation does not result in
CC the constitutive expression of a gene under the control of the phoP
CC regulatory region. The vector sequence encoding this sequence may
CC be used to transform a cell and then the pagC gene product produced.
CC The mutated pagC gene can be used to produce a vaccine comprising a
CC Salmoneilla cell which has been attenuated by the above mutation.
CC These vaccines can be administered orally thus avoiding resistance
CC by patients to vaccination by injection.
SQ Sequence 188 AA;

Query Match 11.8%; Score 108; DB 1; Length 188;
Best Local Similarity 24.2%; Pred. No. 0.0024;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

QY 1 MKTILTLAIVSASSILMSANAISYGSNADAPYVGAKIGQVDAKQINGKNTAYGIYAG 60
DB 1 MKNILSLTVITTSVLVYVNAQADTNFASVGYARAQSKV---QDRKINRGVAVKY----- 54
QY 61 YNPDQNGVEAEFVGSQDKFENAGVSPYKGDYKSGA-----YGT----- 101
DB 54 -----RYEDDSVPSFISLSLYGDRQASGVPEPGIHYHDKFEYKYSGLWGP 102
QY 101 -YRYNFINTPYFAKCKLGIAK-TKVDVTSRNATYTSNK--SDKTSLAGVGVGFPLAVN 156
DB 103 AYRLS-DNFSLYALAGVTVKATREHSTQDQDSFSNKISSRKRTFFAWGAGVOMNPENT 161

Query Match 11.8%; Score 108; DB 1; Length 188;
Best Local Similarity 24.2%; Pred. No. 0.0024;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

QY 1 MKTILTLAIVSASSILMSANAISYGSNADAPYVGAKIGQVDAKQINGKNTAYGIYAG 60
DB 1 MKNILSLTVITTSVLVYVNAQADTNFASVGYARAQSKV---QDRKINRGVAVKY----- 54
QY 61 YNPDQNGVEAEFVGSQDKFENAGVSPYKGDYKSGA-----YGT----- 101
DB 54 -----RYEDDSVPSFISLSLYGDRQASGVPEPGIHYHDKFEYKYSGLWGP 102
QY 101 -YRYNFINTPYFAKCKLGIAK-TKVDVTSRNATYTSNK--SDKTSLAGVGVGFPLAVN 156
DB 103 AYRLS-DNFSLYALAGVTVKATREHSTQDQDSFSNKISSRKRTFFAWGAGVOMNPENT 161

Query Match 12.3%; Score 112; DB 1; Length 362;
Best Local Similarity 28.0%; Pred. No. 0.0023;
Matches 52; Conservative 25; Mismatches 65; Indels 44; Gaps 11;

QY 5 KTLIAVSASSILMSANAISYGSNADAPYVGAKIGQ-----DAKQING----- 51
DB 3 KTLIAVYAGLAAASVQAAPQENTF-----YAGYKAGOGSFDHCINNNGAIKKGLSSSNY 58
QY 51 KNPAYGIYAGYNEFQNG-----GVEAEFY-----GSDAKFENAGVSPYKGDYKSGAYGT 100
DB 62 QGSFHFHAKAHNFEDGGYAGCYLTETRVYTKASENGSD--NF-----GDITS-----K 106
QY 101 YRYNFINTPYFAKCKLGIAKTKVD-VTSRNATYTS--NKSDKTSLAGG-VGVGFKPLAVN 156
DB 107 YAYVTLGKKAHFGVYKLGRAKTIADGITSABEDKEYGVILNNDYIPTSGWTVGYTFKIDGL 166
QY 157 GVEASY 162
DB 167 VLGANV 172

RESULT 7
R70744
ID R70744 standard; protein; 188 AA.
AC R70744;
DT 14-AUG-1995 (first entry)
DE pagC protein.
KW Salmoneilla; vaccine; attenuation; pagC; phoP regulon.
OS Salmoneilla typhimurium.
PN W09502048-A.
PD 19-JAN-1995.
PF 07-JUL-1994; U07658.
PR 09-JUL-1993; US-090526.
PR 06-JUL-1994; US-271354.
PA (GEHO) GEN HOSPITAL CORP.
PA (HARD) HARVARD COLLEGE.
PI Mekalanos JT, Miller SI;
DR WPI: 95-066894/09.
DR N-PSDB; Q85099.

PT Salmonella vaccines conf. cells attenuated by mutation in a *phoP*
PT regulon - are highly attenuated and very safe, partic. for oral
PT admin., mutation may opt. also be in aromatic amino acid
PT synthesis gene.
PS. Disclosure: Page 114-115; 182pp; English.
CC The *phoP* regulatory region activated gene (*pagC*) of wild-type
CC *Salmonella typhimurium* was cloned and sequenced (085099). The *pagC*
CC protein (R07044) is essential for survival within macrophages and
CC for virulence; mutation of *pagC* resulted in attenuation by at least
CC 1,000-fold.
SQ Sequence 188 AA;

Query Match	11.8%	Score 108;	DB 1;	length 188;
Best Local Similarity	24.28%	Pred. No. 0.0024;		
Matches 50;	Conservative 29;	Mismatches 82;	Indels 46;	Gaps 8

```

QY      1 MKKKTLLVASASSLLMSSANAALISYGNSDADQPYGAKIGQVADKQINKNTNAYGIYAG 60
Db      1 MKNIILSTLVTTTSTVLVYVNAQAQDTNAFISGYARIASQSV--ODEKNIRGVNXYK----- 54

QY      61 YNFDQNGVEAEFPGSDAKFEFNAGVSPVKGDPVKSFGA-----YGT----- 101
Db      54 -----RYEDDSVPYSFISLSLTLYIGDRQSGSVPEEGIHYYHDKFEVKKIGSLAVGP 102

QY      101 -YRYNFIINTPYAGKLGIAIK-TRVDYTSRNATTYSNK--SDKTSIAGGVGVGKPLANY 156
Db      103 AYRLS-LDWFSLYALAGVGTWKATPEKHESTQDGDGDSFNKISSRKTFGAMGAGVQGNPLENI 161

QY      157 GVEASY---NYLSEDAANAISLGAIHLAF 180
Db      162 VVDVGVEGSSNISSTKINGFNVGVGYRF 188

```

W18380	
ID	W18380 standard; Protein; 188 AA.
AC	W18380;
DT	27-AUG-1997 (first entry)
DE	S. typhimurium pagC gene expression product.
KW	Salmonella; vaccine; attenuation; diagnosis; pagC gene; Phop;
KW	typhoid fever; virulence.
OS	Salmonella typhimurium ATCC 10428.
FH	Key
FT	peptide
FT	1..23
FT	/label= Sig-peptide
PN	W09720033-A1.
PD	05-JUN-1997.
PF	27-NOV-1996; U019190.
PR	01-DEC-1995; US-565861.
PA	(GENO) GEN HOSPITAL CORP.
PA	(HARD) HARVARD COLLEGE.
PI	Mekalanos JT, Miller SI;
PI	WPI: 97-310590/28.
DR	N-PSDB; T70212.
PT	Attenuated Salmonella strains containing PhoQ or Phop mutation(s) -
PT	used in vaccines, for protection against Salmonella infection and,
PT	e.g. prevention and treatment of typhoid fever
PS	Disclosure; Fig 3; 176pp; English.
CC	The 188-amino acid expression product (W18380) of the Salmonella
CC	typhimurium pagC gene (T70212) is a membrane protein essential for
CC	S. typhimurium survival within macrophages and for virulence.
CC	It shows sequence similarity to the phage lambda protein Lam and
CC	to the all gene product of Yersinia enterocolitica. Salmonella
CC	pad mutant strains are attenuated for virulence and can be used as
CC	live Salmonella vaccines.
CC	Sequence 188 AA;
CC	Sequence 188 AA;

Query Match	11.8%;	Score 108;	DB 1;	Length 188;
Best Local Similarity	24.28;	Pred. No. 0.0024;		
Matches	50;	Conservative 29;	Mismatches 82;	Indels 46;
				Gaps 8

```

QY 1 MKTLEKTLIAVSASSULLMSANMAAISTYGNSSADAPYVCAKIGQVDAKIOJNKNTATYGIYAG 60
Db 1 MKNNILSTVLTVTTSVLVYVNAQAOTNBAESVYATARYAOSKY - ODFKNIRGVNKKY ----- 54
QY 1 YNFDQNFGEAEFPGSDAKEFNAGVSPVKQVSKFSGA - YGT ----- 101
Db 54 ----- RYEDDSVPSFTSSLSLTLYIGDRQAGSVPEBEGIHMHDKFEVYKIGSLMCP 102
QY 101 - YRYNFINTPYRAGKGLIAK - TKVDYTSRNATYYSNK - SDKTSLAGVGVGFKPPLANY 156
Db 103 AYRLS - DNEFSLYALAGVGTVCATKEHSTQGDJSFNKISSRKTGFAMGAGVQMNPLENI 161
QY 157 GVEASY --- NYLSEDAANAISLGAHLAF 180
Db 162 VVDVGYEGSNISSTFKINGFVNGVYRF 188

```

RESULT	9
R96207	
ID	R96207 standard; Protein; 162 AA.
AC	R96207;
DT	24-AUG-1996 (first entry)
DE	Attachment-Invasion-locus protein.
KW	Attachment-Invasion-locus protein; drug delivery;
KW	gastrointestinal membrane; transcytosis; bioavailability;
KW	fusion protein; erythrocyte; Peyer's patch M-cell.
OS	<i>Yersinia enterocolitica</i> strain 8081c.
FH	Key
FT	peptide
FT	19..34
FT	Location/Qualifiers
FT	/note="Peptide from extracellular loop-1"
FT	58..76
FT	/note="Peptide from extracellular loop-2"
FT	102..119
FT	/note="Peptide from extracellular loop-3"
FT	143..152
FT	/note="Peptide from extracellular loop-4"
FT	

PE 20-OCT-1995; U13749.
 PA 27-OCT-1994; US-331393.
 PR (AMGE-) AMGEN INC.
 PT Habbertfield AD, Jensen-Piippo K;
 DR WPI; 96-251447/25.
 NR N-PSDB; 127536.
 PT Therapeutic delivery system utilizing bacterial invasion protein - is
 PT not readily degraded in the gut, enhances systemic bio-availability
 PT of therapeutic agents
 PS Example 1; Fig 2; 110pp; English.
 CC The sequence represents a bacterial attachment-invasion-locus (Ail)
 CC protein, which may be complexed with a therapeutic agent to
 CC transport the agent across the gastrointestinal membrane barrier by
 CC transcytosis to increase bioavailability 5- to 100-fold. The agent
 CC and bacterial protein may be linked via a degradable peptide
 CC sequence as a fusion protein. The receptor binding region of the
 CC Ail protein involves all or some of the regions from the 4
 CC extracellular loops. This region retains the binding affinity of
 CC the protein, and may be used alone or as part of a fusion protein
 CC for drug delivery. The protein may be fused with maltose binding
 CC protein (R96208) to form a fusion protein (R96209) which may be
 CC purified easily by amylose affinity chromatography. The delivery
 CC system allows improved transport across enterocytes and Peyer's
 CC patch M-cells. The system is not prone to degradation in the gut or
 CC early release of biologically active material, and eliminates the
 CC need for parental administration.
 SEQ Sequence 162 AA;

Query Match	11.7%;	Score 106.5;	DB 1;	Length 162;
Best Local Similarity	30.1%;	Pred. No. 0.0028;		
Matches 55;	Conservative 19;	Mismatches 72;	Indels 37;	Gaps 10;
QY	10	VSASSLLMSANAALSY	---GNSADAPYVGAKTGVDAKQINCKNTPAYGIYAGYCNFDQ	65

```
Db 5 VYAAESSISIGYAOQSHVKNQYTLNDP-----KGFNLKY-----REYELD 46
QY 66 NEGVAEAEVGS-DAKEFNAGSPV-KGDYKSGFAGYGRYNE-INTPFYAKGLGIATK 122
Db 47 NMGVIGSFAYTHQGYDFEYGSNKFHGHGVDDYSV--TWGPFSEFRINEYSLYGLGAHAK 104
QY 123 VDVTSRNATTY--SNKSDKTSIAGGVGFKPLANVGVEASTNYLSEDAHAIS---LGAH 177
Db 105 V-----KASVFDESISAKSTMAYGAGVOFNPLPFVVIDASYETSKLDSIVGTWMLGAG 159
QY 178 LAF 180
Db 160 YRF 162

RESULT 10
R85450
ID R85450 standard; Protein; 338 AA.
AC R85450;
DT 15-FEB-1996 (first entry)
DE Nontypable H. influenzae P5 protein.
KW P5 outer membrane protein; vaccine; otitis media; sinusitis;
KW chronic pulmonary obstructive disease.
OS Haemophilus influenzae.
FH Key Location/Qualifiers
FT misc_difference 195
FT /note= "amino acid at position 195 is not
FT /note= "amino acid at position 311
FT /note= "amino acid at position 311 is not
FT /note= "amino acid at position 311 is not
FT misc_difference 311
FT identified in the specification"
FT identified in the specification"
PN EP-680765-A1.
PD 08-NOV-1995.
PR 02-MAY-1995; 302996.
PR 05-MAY-1994; US-210394.
PA (AMCY ) AMERICAN CYANAMID CO.
PI Zlotnick GM.
DR WPI: 95-375029/49.
PT Purified H. influenzae P5 outer membrane protein - used for
PT preventing reducing susceptibility to or treating H. influenzae
PT infections
PS Disclosure: Page 7-8; 16pp: English.
CC Nontypable H. influenzae HI outer membrane protein P5 was isolated
CC by extraction of the outer membrane with detergents and cation-exchange
CC chromatography. P5 (or its peptide fragments) are used in vaccines
CC for prevention of H. influenzae infections implicated in otitis media,
CC sinusitis and chronic pulmonary obstructive disease.
SQ Sequence 338 AA;

Query Match 11.7%; Score 106.5; DB 1; Length 338;
Best Local Similarity 25.6%; Pred. No. 0.0072;
Matches 50; Conservative 18; Mismatches 64; Indels 63; Gaps 11;

QY 35 YVGAKIGQ---VDKQING-----KNT-ANGIYGYNF--DQNGVEAE 72
Db 8 YAGYKAGGSGFHDGINNGAIKEDSIDTLGYGYRNRFTFYGVGGYDILQNDNGLAAE 67
QY 73 FVGSQ-----AKEFNAGVS-PYKGVKSGFAGYGRYNE-INTPFYAKGLGI 117
Db 68 -LGLDNEFRVAFRAEGTKKAHTNHGAHLSLKGSYEVLGDLDY-----GKAG 114
QY 118 IAKTRVD---VTSRNATTYSNKSDKTSIAGG---VGVGFKPLANVGVEAST---NY 164
Db 115 VALVRSDYKFYEAPNSTDAKKGHTHARASGLFAVGAIFYAVLPDLAVLREYQQLRVGKY 174
QY 165 LSEDAHAISLGAHIA 179
Db 175 RPODKNAPSINPNTA 189

RESULT 11
R96210
```

```
ID R96210 standard; Protein; 568 AA.
AC R96210;
DT 24-AUG-1996 (first entry)
DE AIL protein-maltose binding protein fusion protein.
KW Attachment-invasion-locus protein; maltose binding protein;
KW fusion protein; affinity tail; amylose; affinity chromatography;
KW purification; drug delivery; gastrointestinal membrane;
KW transcytosis; bioavailability; enterocyte; Peyer's patch M-cell.
OS Chimeric; Synthetic;.
FH Key Location/Qualifiers
FT region 1..392
FT /note= "Maltose binding protein"
FT region 393..568
FT /note= "Attachment-invasion-locus protein"
FT peptide 425..440
FT /note= "Peptide from extracellular loop-1"
FT peptide 464..482
FT /note= "Peptide from extracellular loop-2"
FT peptide 508..525
FT /note= "Peptide from extracellular loop-3"
FT peptide 551..558
FT /note= "Peptide from extracellular loop-4"
PN M09613250-A1.
PD 09-MAY-1996.
PR 20-OCT-1995; U13749.
PR 27-OCT-1994; US-331393.
PA (AMGE-) AMGEN INC.
PI Habberfield AD, Jensen-Pippo K;
DR WPI: 96-251447/25.
PT Therapeutic delivery system utilising bacterial invasion protein - 1s
PT not readily degraded in the gut, enhances systemic bio-availability
PT of therapeutic agents
PS Example 3; Fig 15; 110pp; English.
CC The sequence represents a fusion protein of a Yersinia
CC enterocolitica attachment-invasion-locus (AIL) protein (R96207)
CC with maltose binding protein (MBP, R86208). The MBP N-terminal
CC portion of the fusion protein acts as an affinity tail, allowing
CC efficient recombinant protein purification by amylose affinity
CC chromatography. The receptor binding region of the AIL protein
CC involves all or some of the regions from the 4 extracellular loops.
CC This region retains the binding affinity of the protein, and may be
CC used alone or as part of a fusion protein for drug delivery. The
CC bacterial AIL protein may be complexed with a therapeutic agent to
CC transport the agent across the gastrointestinal membrane barrier by
CC transcytosis to increase bioavailability 5- to 100-fold. The
CC delivery system allows improved transport across enterocytes and
CC Peyer's patch M-cells. The system is not prone to degradation in the
CC gut or early release of biologically active material, and
CC eliminates the need for parenteral administration.
SQ Sequence 568 AA;

Query Match 11.7%; Score 106.5; DB 1; Length 568;
Best Local Similarity 30.1%; Pred. No. 0.014;
Matches 55; Conservative 19; Mismatches 72; Indels 37; Gaps 10;

QY 10 VASSILLASANAIAISY-----GNSADAPGYGAKIGOVDAKQINGKNTFYGIAGYNFQD 65
Db 411 VYAAESSISIGYAOQSHVKNQYTLNDP-----KGFNLKY-----REYELD 452
QY 66 NEGVAEAEVGS-DAKEFNAGSPV-KGDYKSGFAGYGRYNE-INTPFYAKGLGIATK 122
Db 453 NMGVIGSFAYTHQGYDFEYGSNKFHGHGVDDYSV--TWGPFSEFRINEYSLYGLGAHAK 510
QY 123 VDVTSRNATTY--SNKSDKTSIAGGVGFKPLANVGVEASTNYLSEDAHAIS---LGAH 177
Db 511 V-----KASVFDESISAKSTMAYGAGVOFNPLPFVVIDASYETSKLDSIVGTWMLGAG 565
QY 178 LAF 180
Db 566 YRF 568
```


1
1
1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 12:18:21; Search time 27.94 Seconds

(Without alignments)

93.180 Million cell updates/sec

Title: US-09-164-714-7

Perfect score: 912
Sequence: 1 MTKTLVAVAGLAASVADAPQENFTF---YAGVRAAGSGSFHDGJNNNGAIKKGSLSSNY 180

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/PCRTUS.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	12.4	359	1	US-08-457-997B-2
2	112.5	12.3	361	4	PCT-US94-08326-8
3	108	11.8	188	1	US-08-286-767-3
4	106.5	11.7	568	4	PCT-US95-13749-5
5	105.5	11.6	338	1	US-08-210-394-1
6	103	11.3	188	1	US-08-486-715-5
7	103	11.3	188	1	US-08-486-719-5
8	103	11.3	188	1	US-08-476-100-5
9	103	11.3	188	3	US-08-475-749-5
10	103	11.3	188	3	US-08-475-749-5
11	100	11.0	1287	1	US-08-200-232-2
12	100	11.0	1287	4	PCT-US95-02219-2
13	100	11.0	1287	4	PCT-US95-02219A-2
14	87.5	9.6	363	4	PCT-US94-08326-10
15	83.5	9.2	455	2	US-08-472-172-4
16	83	9.1	759	2	US-08-450-351-2
17	83	9.1	759	2	US-08-450-351-2
18	80.5	8.8	342	4	PCT-US94-08326-12
19	80.5	8.8	342	4	PCT-US94-08326-14
20	80.5	8.8	1612	1	US-08-169-927-2
21	75.5	8.3	663	1	US-08-765-081-7
22	74.5	8.2	792	3	US-08-433-522A-8
23	74.5	8.2	1394	4	PCT-US95-10661A-2
24	74	8.1	703	4	PCT-US95-06994-9
25	73.5	8.1	340	2	US-08-355-844-1
26	73.5	8.1	340	4	PCT-US95-16126-1
27	72.5	7.9	205	1	US-08-277-231A-5
28	72.5	7.9	205	2	US-08-473-750-8
29	72.5	7.9	205	2	US-08-477-326-8

30	72	7.9	251	1	US-08-209-747-8	Sequence 8, Appl1
31	72	7.9	251	1	US-08-458-298-8	Sequence 8, Appl1
32	72	7.9	459	2	US-08-472-172-2	Sequence 2, Appl1
33	71.5	7.8	1912	1	US-08-409-995-4	Sequence 4, Appl1
34	71	7.8	584	1	US-08-313-288B-17	Sequence 17, Appl1
35	71	7.8	593	1	US-07-961-522-4	Sequence 4, Appl1
36	71	7.8	593	1	US-08-217-438-4	Sequence 4, Appl1
37	71	7.8	593	1	US-08-217-438-5	Sequence 4, Appl1
38	71	7.8	593	1	US-08-487-890A-100	Sequence 100, App
39	71	7.8	593	1	US-08-321-978-4	Sequence 4, Appl1
40	71	7.8	593	2	US-08-710-584-4	Sequence 100, App
41	71	7.8	593	2	US-08-478-435-100	Sequence 100, App
42	71	7.8	593	2	US-08-337-483-100	Sequence 100, App
43	71	7.8	593	2	US-08-478-373-100	Sequence 100, App
44	71	7.8	593	3	US-08-474-671-100	Sequence 100, App
45	71	7.8	593	3	US-08-474-671-100	Sequence 100, App

ALIGNMENTS

```
RESULT 1
US-08-457-997B-2
: Sequence 2, Application US/08457997B
: Patent No. 5766608
: GENERAL INFORMATION:
: APPLICANT: Kolattukudy, P. E.
: TITLE OF INVENTION: Otitis Media Vaccine
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESS: Calfee, Halter and Griswold
: STREET: Suite 1800 800 Superior Avenue
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: U.S.A.
: ZIP: 44114-2688
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457, 997B
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Goltick, Mary E.
: REGISTRATION NUMBER: 34,829
: REFERENCE/DOCKET NUMBER: 22727/00102
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (216) 622-8458
: TELEFAX: (216) 241-0816
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 359 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-457-997B-2
```

Query Match 12.4%; Score 113; DB 1; Length 359;
Best Local Similarity 28.2%; Pred.No. 0.00011;
Matches 55; Conservative 18; Mismatches 80; Indels 42; Gaps 11;

```
OY 5 KTLVAVAGLAASVADAPQENFTF---YAGVRAAGSGSFHDGJNNNGAIKKGSLSSNY 51
   ||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
OY 51 ---KMT-AYGIAGYVNF---DONFGVEAEFVGS---AKEFNAGVSPVKGVDVSGFAY--- 99
   ||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
DB 3 KTLVAVAGLAASVADAPQENFTF---YAGVRAAGSGSFHDGJNNNGAIKKGSLSSNY 58
DB 59 GYRRTFTFYGVGGVQIINQDFGLAALGLGYPDFGRAKTRKGRKRAKH---TNGATYSTL 116
```

QY	99	-GPRXNFMPTFAKCKLIATKVD--VTSRMATTSNKSODKSLAGC---	VGCGEK	151
	117	KGSE---VLDDGYKKAQALVRSYKFEEDANGTGRHKKGRHTARASGLFAVGAET		173
Db	152	PLANVGEASTINILS	166	
QY	174	VLPELAVREYQWLT	188	

```

RESULT      2
PCT-US94-08326-8
/ Sequence 8, Application PC/US94408326
/ GENERAL INFORMATION:
/ APPLICANT: North American Vaccine, Inc.
/ APPLICANT: 12103 Indian Creek Court
/ APPLICANT: Beltsville, MD 20705
/ APPLICANT: Pullen, Jeffrey K.
/ APPLICANT: Soper, Thomas S.
/ APPLICANT: Liang, Shu-Mei
/ TITLE OF INVENTION: A Method For The High Level
/ TITLE OF INVENTION: Expression,
/ TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
/ TITLE OF INVENTION: Protein
/ TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Sterne, Kessler, Goldstein & Fox
/ STREET: 1100 New York Avenue, Suite 600
/ City: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20005-3934
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US94/08326
/ FILING DATE: Herewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/096,181
/ FILING DATE: 23-JULY-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Esmond, Robert W.
/ REFERENCE/DOCKET NUMBER: 1438.001PC01
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 371-2600
/ TELEFAX: (202) 371-2540
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 361 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
PCT-US94-08326-8

```

Query Match	12.3%	Score 112.5;	DB 4;	Length 361;
Best Local Similarity	28.1%	Pred. No. 0.00013;		
Matches 52; Conservative	25;	Mismatches 65;	Indels 43;	Gaps 11;

[illegible]

Db 107 AYVVLGNKAEEVYLGNAKTIADGITSAEDEKEIGVLNNSDIPTSGATVGYTFKIGDGLV 166

QY 158 VEASY 162
: : :
Db 167 LGANT 171

```

RESULT 3
US-08-286-767-3
: Sequence 3, Application US/08286767
: Patent No. 5733760
: GENERAL INFORMATION:
: APPLICANT: Lu, Yichen
: APPLICANT: Miller, Samuel I.
: APPLICANT: Killeen, Kevin
: TITLE OF INVENTION: SALMONELLA VECTORS ENCODING TRUNCATED
: TITLE OF INVENTION: P93C FUSION PROTEIN, METHOD OF MAKING, AND USES THERETO
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: RONALD I. EISENSTEIN; DIKE, BRONSTEIN,
: ADDRESSEE: ROBERTS & CUSHMAN
: STREET: 130 Water Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: US
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/286,767
: FILING DATE: 05-AUG-1994
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Resnick, David S.
: REGISTRATION NUMBER: 34235
: REFERENCE/DOCKET NUMBER: 44420
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 523-3400
: TELEFAX: (617) 523-6400
: TELEX: 200291 STRE UR
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 188 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-286-767-3

```

Query Match	11.8%	Score 108; DB 1;	Length 188;
Best Local Similarity	24.2%	Pred. No. 0.00017;	
Matches 50; Conservative	29;	Mismatches 82;	Indels 46; Gaps 8;

[illegible]

RESULT 4
 PCT-US95-13749-5
 ; Sequence 5, Application PC/TUS9513749
 ; GENERAL INFORMATION:
 ; APPLICANT: Amgen Inc.
 ; TITLE OF INVENTION: COMPOSITIONS FOR INCREASED
 ; BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amgen Inc.
 ; STREET: 1840 Dehavilland Drive
 ; CITY: Thousand Oaks
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 91320-1789
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/13749
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 568 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ;
 ;
 PCT-US95-13749-5

```

Query Match          11.7%; Score 106.5; DB 4; Length 568;
Best Local Similarity 30.1%; Pred. No. 0.0011;
Matches      55; Conservative     19; Mismatches    72; Indels   37; Gaps    10;

Oy      10 VASASSLLAMNANAISY-----GNSADAPFVGAKIGGVDAKQJNGKNTAVGIYAGYNPDQ 65
       | :| :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      411 YVAASSDSISITGMOASHVENGTTLNDP-----KGFMUKY----RYELDD 452

Oy      66 NEGEAEFVGS-DAKEFNAGVSPV-KGDVKSFGAYGTRYRNF-INTPFYAKGLGIATRK 122
       | :| :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      453 NMGVGSGFAHYTHQGVDYFPFGSNKFEGHGDVDYYSV--TMGPSFRINEVSYLGLLGAAHGK 510

Oy      123 VDVTSMNATTY--SNKSDKTSIAGVGYCGFKPLANNVEASYNTLSDDANAIIS---LGAH 177
       | :| :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      511 V-----KASVPDESISASTSMAYGAGVOFNPINFVIDASYETSKLDSIKVGTWMLGAG 565

Oy      178 LAF 180
       |
Db      566 YRF 568

RESULT      5
US-08-210-394-1
; Sequence 1, Application US/08210394
; Patent No. 5770213
; GENERAL INFORMATION:
; APPLICANT: Zlotnick Dr., Gary W.
; TITLE OF INVENTION: Purified No. 5770213typable Haemophilus
; TITLE OF INVENTION: Influenzae P5 Protein as a Vaccine for No. 5770213typable
; TITLE OF INVENTION: Haemophilus Influenzae Strain
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: US
; ZIP: 07470-8426

COMPUTER READABLE FORM:
```

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: Patentin Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/210,394
7 FILING DATE: 07-MAR-1994
8 CLASSIFICATION: 424
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Harrington, James J
11 REFERENCE/DOCKET NUMBER: 32,144
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: 201/831-3246
14 TELEFAX: 201/831-3305
15 INFORMATION FOR SEQ ID NO: 1:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 338 amino acids
18 TYPE: amino acid
19 STRANDEDNESS: unknown
20 TOPOLOGY: unknown
21 MOLECULE TYPE: protein
22 US-08-210-394-1

```

```

Query Match      11.6%; Score 105.5; DB 1; Length 338;
Best Local Similarity 25.6%; Pred. No. 0.00071;
Matches 50; Conservative 17; Mismatches 65; Indels 63; Gaps 11.

QY 35 YVGATIGQ---VDANQING-----KNT-AYGIYAGYNF--DQNGVEAE 72
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 8 YAGVAGAGGQSFHDGINNCAIKEDSIDLTGIGYRRNFTFYGVFGYQILNODNGLAE 67
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 73 FVGSD-----AKEFNAGVS-PYKGDVKSFGAGTYRYNFINTPEYAKGKLG 117
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 68 -LGYNFGRVKRRARGKTKAKHTNINGAHLSTLGSVEVDLDGLVY-----GKAG 114
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 IAKTVD---VTSRNATYYSNKSCKDTSLAG---VGVGKPLANNGVSEASY-----NY 164
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 115 VALVRSDYKFFYAPAPSTRUKKKGHTARASGLFAVGAEEYAVLPFLAVRLEYOQLTRVGRY 174
      | : | | | : | : | : | : | : | : | : | : | : | : | : |
QY 165 LSEDANATISLAHLLA 179
      | : | | | : | : | : | : | : | : | : | : | : | : | : |
DB 175 RPQDKNAPSINPNTA 189
      | : | | | : | : | : | : | : | : | : | : | : | : | : |

RESULT 6
US-08-486-715-5
; Sequence 5, Application US/08486715
; Patent No. 5674724
; GENERAL INFORMATION:
; APPLICANT: MILLER, Samuel I.
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson P. C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,715
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/090,526
; FILING DATE: 09-JUL-1993
; PRIOR APPLICATION DATA:

```

```

1 APPLICATION NUMBER: 07/629,602
2 FILING DATE: 18-DEC-1990
3 ATTORNEY/AGENT INFORMATION:
4 NAME: Clark, Paul T.
5 REGISTRATION NUMBER: 30,162
6 REFERENCE/DOCKET NUMBER: 00786/192005
7 TELECOMMUNICATION INFORMATION:
8 TELEPHONE: (617) 542-5070
9 TELEFAX: (617) 542-8906
10 TELEX: 200154
11 INFORMATION FOR SEQ ID NO: 5:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 188 amino acids
14 TYPE: amino acid
15 STRANDEDNESS: not relevant
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18 OS-08-486-715-5

```

Query Match	11.3%	Score 103;	DB 1;	Length 188;
Best Local Similarity	27.0%	Pred. No.	0 00059;	
Matches 53;	Conservative 28;	Mismatches 91;	Indels 24;	Gaps 10;

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

?      REGISTRATION NUMBER: 30,162
?      REFERENCE/DOCKET NUMBER: 00766/192002
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (617) 542-5070
?      TELEFAX: (617) 542-8906
?      TELEX: 200154
?      INFORMATION FOR SEQ ID NO: 5:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 188 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: not relevant
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
?
US-08-486-719-5

```

Query Match	11.3%	Score 103;	DB 1;	length 188;
Best Local Similarity	27.0%	Pred. No.	0	00059
Matches	53;	Conservative	28;	Mismatches 91;
			Indels	24;
			Gaps	10

[illegible]

RESULT 8
 US-08-476-100-5
 ; Sequence 5, Application US/08476100
 Patent No. 5731196
 GENERAL INFORMATION:
 APPLICANT: Miller, Samuel I., III
 APPLICANT: Mekalanos, John J.
 TITLE OF INVENTION: SALMONELLA VIRULENCE GENES
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P. C.
 STREET: 225 Franklin Street
 City: Boston
 STATE: Massachusetts
 COUNTRY: U S A
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 502 or 55SX
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: Wordperfect (Version 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/476.100
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/090,526
 FILING DATE: 09-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/629,602
 FILING DATE: 18-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Clark, Paul T.
 REGISTRATION NUMBER: 30,162
 REFERENCE/DOCKET NUMBER: 00786/192003
 TELECOMMUNICATION INFORMATION:

```

? TELEPHONE: (617) 542-5070
? TELEFAX: (617) 542-8906
? TELLEX: 200154
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 188 amino acids
? TYPE: amino acid
? STRANDEDNESS: not relevant
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? OS=08-476-100-5

```

Query Match	11.3%	Score	103;	DB	1;	Length	168;
Best Local Similarity	27.0%	Pred. No.	0.00059;				
Matches	53;	Mismatches	51;	Indels	24;	Gaps	10;
		Conservative	28;				

```

QY 61 YNFDQ--NFGVEAEFVSGDAKEFNAGVSPY-----KGDKV--SFGAYGTYRYNFINTPP 110
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 55 YEDSDPVSFISLSLYELGD--RQASGVSPEBGITHYHDKFEVKKXSLMGPAVRLS--DNFSL 112
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
QY 111 YAKGLCLIAK--TKVDYTSRNAATYSNF--SDKTISLACGVGVGERPLANVGVEASY---NY 164
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 113 YALAGVTVKATPEKHSTQDODSPFSNRKTSRKTFGAMGAGVQMPLENIVVDVGYECSNI 172
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
QY 165 LSEDAANAISLGAHLAF 180
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 173 SSTKINGFNVGVGYRF 188
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

```

9
US-08-475-749-5 / Sequence 5, Application US/08475749
Patent No. 6010901
GENERAL INFORMATION:
APPLICANT: Miller III, Samuel I.
TITLE OF INVENTION: SALMONELLA VIRULENCE GENES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P. C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,749
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/090,526
FILING DATE: 09-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/629,602
FILING DATE: 18-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/192004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

```

; INFORMATION FOR SEQ ID NO: 5:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 188 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: not relevant
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
US-08-475-749-5

```

Query Match	11.3%	Score 103;	DB 3;	Length 188;
Best Local Similarity	27.0%;	Pred. No. 0.00059;		
Matches 53; Conservative	28;	Mismatches 91;	Indels 24;	Gaps 10

[illegible]

US-RSULT 10
Sequence 5, Application US/08475749
Patent No. 6019001
GENERAL INFORMATION:
APPLICANT: Miller III, Samuel I.
APPLICANT: Mekalanos, John L.
TITLE OF INVENTION: SALMONELLA VIRULENCE GENESSES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P. C.
STREET: 225 Franklin Street
City: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
Zip: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,749
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/090,526
FILING DATE: 09-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/629,602
FILING DATE: 18-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/192004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids

;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-475-749-5

Query Match 11.3%; Score 103; DB 3; Length 188;
Best Local Similarity 27.0%; Pred. No. 0.00059;
Matches 53; Conservative 28; Mismatches 91; Indels 24; Gaps 10;

QY 1 MKTTLTLLASVSSLLSANAISYGSNADAPYVGAKIGVDKQINRKTATGITYAG 60
DB 1 MKNITLSTLVITTSVLVNVNAQADTNAPSVGARYAQSKV--QDFKNIRGVNVKTY---R 54
QY 61 YNFDQ--NFGVEAEFVGSDAKEFNAGVSPV-----KGDVK--SFGAYGTYYRNFINTPEF 110
DB 55 YEDDSPVPSPFISLSLYLGD-RQASGSVEPEGIHYHDKFEVKKGLSLMGVARYLALD-DFEFL 112
QY 111 YAKGLGLIAK-TKVDVTSRNATTYSNK--SDKTSIAGVGVGFKPLANGVBAASY---NY 164
DB 113 YALAGVGTWKATFKFKEHSYQDDSFNSKTSRKTGFAMGAGVQMNPLENTIVDVYEGSNI 172
QY 165 LSEDAANAISLGAHLAF 180
DB 173 SSTKINGFNVGVGRF 188

RESULT 11

US-08-200-232-2
Sequence 2, Application US/08200232
Patent No. 5721349

GENERAL INFORMATION:
APPLICANT: Cover, Timothy L.
TITLE OF INVENTION: Blaser, Martin J.
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
TITLE OF INVENTION: AND RELATED METHODS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NEEDLE & ROSENBERG P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200.232
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spiralt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-200-232-2

Query Match 11.0%; Score 100; DB 1; Length 1287;
Best Local Similarity 24.8%; Pred. No. 0.018;
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;

QY 9 AVSASSLLAMSANNAISYGSNADAPYVGAKIGVDA-----KQIN----- 50
DB 1025 AIGGAS-LNNGGNASL-YETSAAGVDAYLN---GQVEAIVGGFGSYGSFNNQANSLSNG 1079
QY 50 GKNTAYGIYAGYNFDON-FGVBAE-FVGSDAKEFNAGVSPVKGDKSF-----GA 97
DB 1080 ANNTNFGVYSRIFANQHEDEFAOGALGSDOSLNFKSLALDLQSYNYLAYSAATRAS 1139
QY 98 YGTYYRNFINTPEFVAKGKIGIAKTVDTSRNATTYSNKSDDKTSIAGVGVGFKPLANVG 157
DB 1140 YG-YDPAFERNALVLKPSGVSYNHLGSTNEK-----SNTNRVALSNGSSOHLNANAN 1194
QY 158 VEASYN-----LSE-----DANAISL 174
DB 1195 YEARYYGGTTSYFYMNAGVLQEFANRGSNAVSL 1228

RESULT 12

PCT-US95-02219-2
Sequence 2, Application PC/TUS9502219
GENERAL INFORMATION:

APPLICANT: Cover, Timothy L.
TITLE OF INVENTION: Blaser, Martin J.
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
TITLE OF INVENTION: AND RELATED METHODS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NEEDLE & ROSENBERG P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02219
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Spiralt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1287 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02219-2

Query Match 11.0%; Score 100; DB 4; Length 1287;
Best Local Similarity 24.8%; Pred. No. 0.018;
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;

QY 9 AVSASSLLAMSANNAISYGSNADAPYVGAKIGVDA-----KQIN----- 50
DB 1025 AIGGAS-LNNGGNASL-YETSAAGVDAYLN---GQVEAIVGGFGSYGSFNNQANSLSNG 1079
QY 50 GKNTAYGIYAGYNFDON-FGVBAE-FVGSDAKEFNAGVSPVKGDKSF-----GA 97
DB 1080 ANNTNFGVYSRIFANQHEDEFAOGALGSDOSLNFKSLALDLQSYNYLAYSAATRAS 1139
QY 98 YGTYYRNFINTPEFVAKGKIGIAKTVDTSRNATTYSNKSDDKTSIAGVGVGFKPLANVG 157

Db 1140 YG-YDFAFERNALVKPSGVSYNHLGSTNFK-----SNSTNKVALNSGSSQHLFNASAN 1194
QY 158 VEASYN-----LSE-----DANAISL 174
Db 1195 VEARYYGDTSYFYMAGVLOEFANFGSSNAVSL 1228

RESULT 13

PCT-US95-02219A-2
; Sequence 2, Application PC/TUS9502219A
; GENERAL INFORMATION:
; APPLICANT: Cover, Timothy L.
; APPLICANT: Tumuru, Murali KR
; APPLICANT: Cao, Ping
; APPLICANT: Thompson, Stuart A.
; APPLICANT: Blaser, Martin J.
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
; TITLE OF INVENTION: AND THE RELATED METHODS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02219A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-02219A-2

Query Match 11.0%; Score 100; DB 4; Length 1287;
Best Local Similarity 24.8%; Pred. No. 0.018;
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;

QY 9 AVASSSLANSANAIAISGSAQAQPYGAKIGQYDA-----KQIN----- 50
Db 1025 AIGAS--LNNGNASL-YGTSAGVDAYLN--GQVEAIVGFGSYGYSSFNQANSLSNG 1079
QY 50 GKNTAYCIYAGYNDON-FGVEAE-FVSGDAKEFNAGVSPYKGVKSP-----GA 97
Db 1080 ANNTNFYVIRFANQHEFDEFAOGALGSDOSSLNFKSALIRLINOYNTIYATSNATRAS 1139
QY 98 YGTYRYNFINTPYAKGKLGIATKTVDTSRNATYTSNKSQDTSLAGGVGFPLANVG 157
Db 1140 YG-YDFAFERNALVKPSGVSYNHLGSTNFK-----SNSTNKVALNSGSSQHLFNASAN 1194
QY 158 VEASYN-----LSE-----DANAISL 174
Db 1195 VEARYYGDTSYFYMAGVLOEFANFGSSNAVSL 1228

RESULT 14

PCT-US94-08326-10
; Sequence 10, Application PC/TUS9408326
; GENERAL INFORMATION:
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Wei
; TITLE OF INVENTION: A Method For the High Level
; TITLE OF INVENTION: Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08326
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,181
; FILING DATE: 23-JULY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REFERENCE/DOCKET NUMBER: 1438.001PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-08326-10

Query Match 9.6%; Score 87.5; DB 4; Length 363;
Best Local Similarity 26.7%; Pred. No. 0.074;
Matches 47; Conservative 24; Mismatches 62; Indels 43; Gaps 11;

QY 14 SLMSANAAIAISYGSNAQAQPYGAKIGQY-----DAKQING-----KNTAYGIYAG 60
Db 14 SLPVSSDPAYVYNNEG-TNVELGRLSLIEQSNSTYVNDQOQALRNGSSRHIAAT 72
QY 61 YNFDONF-----GVEAEFY-----GSDAKFENAGVSPYKGVKSPGAYGTYRYNFINTPF 110
Db 73 HNFQGGFQAQGLYLRFRFTKASENGSD--NF-----GDITS-----KAYVYLGKKA 117
QY 111 YAKGLGIATKTVQD-VISRNATYTS--NKSQDTSLAGG-VGVGFPLANVGVEASY 162
Db 118 FGEVYKLAGRAKTIADGITSADKEYGVNLNNSDIPTSGWTGVGTGKIDGLVIGANY 173

RESULT 15
US-08-472-172-4
; Sequence 4, Application US/08472172
; Patent No. 5985288
; GENERAL INFORMATION:
; APPLICANT: Munson, Jr., Robert S
; APPLICANT: Grass, Susan

APPLICANT: Chong, Pele Y Y
APPLICANT: Fahim, Raifat
APPLICANT: Sia, Charles D Y
APPLICANT: McVerity, Patrick
APPLICANT: Klein, Michel
TITLE OF INVENTION: Outer Membrane Protein P1 and Peptides
TITLE OF INVENTION: of Haemophilus Influenzae Type B
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R6
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,172
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/849,411
FILING DATE: 07-JUL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-471
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
TELEX: 063-24567 SIMBAS
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-472-172-4

Query Match 9.2%; Score 83.5; DB 2; Length 455;
Best Local Similarity 26.0%; Pred. No. 0.28;
Matches 44; Conservative 33; Mismatches 63; Indels 29; Gaps 11;
QY 13 SSLAMSAANAISYGNADAPYGAIGOV-DAKQNGKNTAYGIYAGYNFQNGVEA 71
DB 86 ASIATTKMNSA-KYGSASERNVPGAFVNLVFAVPVNDK---FALGAGMNV--NFGLS 139
QY 72 EFVGSDAKEFNAGVSPYKGDVK--SRGAYGTRYNFINTPFYAKGLGT---AKTKVDY 125
DB 140 EYDVS---YDAGIFGKTDLTAINLNLGAYRV---TEGLSLG-LGVNAVYAKAQVE- 190
QY 126 TSRNATTYNSKSDKSLAGVGVEFKPLANVGVEASYNLSEDAANISL 174
DB 190 --RNAGIIANSVNDIOVTALSYLAPLKLGL---DQNLPSKDKSVVSL 232

Search completed: May 20, 2000, 12:18:25
Job time: 6110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 12:19:13 ; Search time 43.2 Seconds
(without alignments) 244.291 Million cell updates/sec

Title: US-09-164-714-7

Perfect score: 912
Sequence: 1 MKTLTILAVSASSILAMSA.....SYNYISEDANAISLAHLAF 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123.5	13.5	178	F64124	opacity protein ho
2	123	13.5	178	A35123	all protein precur
3	112.5	12.3	359	D64050	major outer membra
4	112.5	12.3	361	A30542	major outer membra
5	109.5	12.0	264	154668	heat resistant agg
6	108	11.8	188	A39185	virulence protein
7	108	11.8	365	MMBP2	outer membrane por
8	108	11.8	365	1	outer membrane por
9	103	11.3	371	MMECNC	major outer membra
10	103	11.3	371	S68072	major outer membra
11	102.5	11.2	213	A56152	major 25k outer me
12	101	11.1	386	S68062	major outer membra
13	101	11.1	386	S68063	major outer membra
14	101	11.1	386	S68064	major outer membra
15	100	11.0	1287	B53739	vacuolating cyto
16	100	11.0	1291	S44983	vacuolating cyto
17	99.5	10.9	172	A39189	outer membrane pro
18	98.5	10.8	215	JC4087	accessory coloniza
19	96	10.5	385	S68066	major outer membra
20	96	10.5	385	S68067	major outer membra
21	96	10.5	385	S68068	major outer membra
22	96	10.5	385	S68070	major outer membra
23	96	10.5	385	S68071	major outer membra
24	95.5	10.5	1290	G64630	vacuolating cyto
25	95	10.4	211	S47347	outer membrane pro
26	95	10.4	575	S41310	flagellin A - Camp
27	95	10.4	575	140615	flagellin protein
28	95	10.4	1288	E71884	vacuolating cyto
29	90	9.9	185	A43309	outer membrane pro
30	90	9.9	196	S70957	otnd protein - Vib

31	90	9.9	575	2	S41311	flagellin B - Camp
32	89.5	9.8	1585	2	B69948	phage-related prot
33	89	9.8	171	2	I55173	outer membrane pro
34	89	9.8	359	2	S07298	outer membrane pro
35	88	9.6	346	1	MMECA	outer membrane pro
36	86	9.4	394	2	S34263	outer membrane por
37	86	9.4	623	2	164034	hypothetical prote
38	86	9.4	651	2	E71904	probable outer mem
39	85.5	9.4	356	2	A55016	lysosomal membrane
40	85.5	9.4	483	2	A71958	outer membrane pro
41	85.5	9.4	711	2	A64523	outer membrane pro
42	85.5	9.4	2340	2	B71704	cell surface antiq
43	85	9.3	503	2	T35053	probable solute-bi
44	85	9.3	638	2	B71904	probable outer mem
45	85	9.3	691	2	C64548	outer membrane pro

ALIGNMENTS

RESULT 1
F64124
opacity protein homolog H11457 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: F64124
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L., McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: F64124
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1178 <TIGR>
A:Cross-references: GB:U32823; GB:L42023; NID:g1574281; PIDN:AAC23104.1; PID:g1574294

Query Match 13.5% Score 123.5; DB 2; Length 178;
Best Local Similarity 27.5%; Pred. No. 0.0014;
Matches 52; Conservative 30; Mismatches 82; Indels 25; Gaps 8;
QY 4 LKTLAVSASSILAMSAANAISYGSADAPYGAIG--QYDAKQINGKNT---AYGTY 58
Db 1 MKKLIVMLFTIAL-----SAQAQWYVGDLGASKRIDITHVSSNPSTQRTS 50
QY 59 AGYNFDQNGVEAEFVGSDAKEFN-AGVSPYKGVKDYKSGFANGTYRYNFINPFYAKKLG 117
Db 51 VGAYADKNRFLAVDYTNYGKVTANADVDVSLKCKSLGILTFYDFDLADFPYV--GVR 108
QY 118 IAKTVDTNRNATYYSKSDKSLAGV---GYGFRLANVGVAS--VYLSIEDANA 171
Db 109 VSTNGADVTA-NARYRYLEAFETTRIGIGALAGVQKLTQDVAINTNIEYRLASNVSD 167
QY 172 ISGAHLAF 180
Db 168 VGKAGLRF 176
RESULT 2
A35123
all protein precursor - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 08-Oct-1999
C:Accession: A35123
R:Miller, V.L.; Bilska, J.B.; Falkow, S.
J. Bacteriol. 172, 1062-1069, 1990
A:Title: Nucleotide sequence of the Yersinia enterocolitica All gene and characterize
A:Reference number: A35123; MUID:90130261
A:Accession: A35123
A:Status: preliminary

heat resistant agglutinin 1 - Escherichia coli
C:Species: Escherichia coli
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Oct-1999
C:Accession: I54668; I69133; I69134
R:Lutwyche, P.; Rupp, R.; Cavanagh, J.; Warren, R.A.; Brooks, D.E.
Infect. Immun. 62, 5020-5026, 1994
A:Title: Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglut
A:Reference number: I54668; MUID:95012721
A:Accession: I54668
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-264 <RES>
A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AC13751.1; PID:g463911
A:Accession: I69133
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 15-264 <RES>
A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AC13752.1; PID:g463912
A:Accession: I69134
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 18-264 <RES>
A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AC13753.1; PID:g463913

Query Match 12.0%; Score 109.5; DB 2; Length 264;
Best Local Similarity 21.0%; Pred. No. 0.033;
Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;

Qy 1 MKTKTLAVSASSLLMSANAIS-----YGSNDAQP 34
Db 15 MIENKXIVASALMAGSFSTQALADESKTGYTGKAGASVMSIADORELSGNGEELSK 74
Qy 35 YVAKIGQVADAKQINGKNFAY--GIYAGYFQDNFQV---EAEFV--GSDAKFENAGVS 86
Db 75 YKGG-----DGHDTVFSGGIAAGYDFYPOFSIPVTELEFYARGKADSKYNDKD 124
Qy 87 PVKG-----DYKS-----FGAYGYTRYFINTPTFAKRLGIAR-----TKDV-- 126
Db 125 SMSGGYWRDLDKNEVSVNTLMNLAYDFRINDSAFTPWVSAG--IGYAKRHHOKTGTISTWD 183
Qy 126 -----TSRNATTYSNKSDDKSLAGVGVEFKPLANVGVEASYNLSEDAANIS 173
Db 184 YGYSSGRESLSRSGSADNFAMSLGAGRYDYTPDIALDLSTRLDAGDSSVS 236

RESULT 6
A39185
virulence protein pagC precursor - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 08-Oct-1999
C:Accession: A39185
R:Pulkkinen, W.S.; Miller, S.I.
J. Bacteriol. 173, 86-93, 1991
A:Title: A Salmonella typhimurium virulence protein is similar to a Yersinia enterocolit
A:Reference number: A39185; MUID:91100323
A:Accession: A39185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <PUL>
A:Cross-references: GB:M55546; NID:g154232; PIDN:AAA27179.1; PID:g154233

Query Match 11.8%; Score 108; DB 2; Length 188;
Best Local Similarity 24.2%; Pred. No. 0.03;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

Qy 1 MKTKTLAVSASSLLMSANAISYGSADAPRYGAKIGQVADAKQINGKNFATGIRAG 60
Db 1 MKNILLSTLVITTSVLVNVAAQDTNAPSVGARYAOSKV--QDFKTRGVNVRT----- 54
Qy 61 YNPDNGVEAEFYGSDAKFENAGVSPYKGVKSGA-----YGF----- 101

Db 54 -----RYEDDSPVFSISLSLYGDRQAGSVEPEGIHYHDKFEVYKSLWGP 102
Qy 101 -YRYNFITPPYAKKGLGIAR-TKYDVTSRNATTYSNR--SKTSLAGVGVEFKPLANV 156
Db 103 AYRRS-DNFSLYALAGVGTAKATFKESHSTODDSFNSNISKTKGFAMGAGVQNNPLENI 161
Qy 157 GVEASY---NYLSEDAANISGAHLAF 180
Db 162 VVDVYEGSNISSTKINGFNWGVGYRF 188

RESULT 7
MMBP2
outer membrane porin lc precursor - phage PA2
C:Species: phage PA2
A:Note: host Escherichia coli
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 02-Jul-1998
C:Accession: D25647
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.
J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the lc and mmpC outer membrane porin protein genes of lambdaoid
A:Reference number: A25647; MUID:86304457
A:Accession: D25647
A:Molecule type: DNA
A:Residues: 1-365 <BLA>
C:Genetics:
A:Gene: lc
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-23/domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin lc #status predicted <MAT>

Query Match 11.8%; Score 108; DB 1; Length 365;
Best Local Similarity 26.6%; Pred. No. 0.064;
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;

Qy 1 MKTKTL-AYSASSLLMSANAISY-----GNSDAQRYGAK 39
Db 1 MKKLVAISAAVASVLMASQAARLYNKDSNKLDTYGVNAKHVSSNDADDGDTTYAR 60
Qy 40 IGOVDKQINGKNFAYGIYAGYFQDNFQVGEAEFVYGSADAKFENAGVSPYKGVKSGAYG 99
Db 61 LQFKGTQINDQLTGFGWE-YEFKGN--RAESGSSKDKTRLAPAGIK-----FGDIG 111
Qy 100 YTRYNFINTPTFAKGL-----GIATKYDV-----TSRN----- 130
Db 112 SIDYGRNKGAVADIGAMVDYLPFEGDPTWTCYDVMGTGTFATRYRNNDPFGVDGLNF 171
Qy 130 ATTYSNKSDDKT---SLAGVGVEFKPLANVGVEASYNLSEDAANISGAHLA 179
Db 172 AAQYQKNDKRDSPFDNYTEGNGDF-----GFSATYEV-----EGFGIGATYA 213

RESULT 8
MMBNC
outer membrane porin mmpC precursor - Escherichia coli cryptic lambdaoid prophage DLP1
C:Species: Escherichia coli
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A25647; S66594; G64787
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.
J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the lc and mmpC outer membrane porin protein genes of lambdaoid
A:Reference number: A25647; MUID:86304457
A:Contents: mutant strain CS384
A:Accession: A25647
A:Molecule type: DNA
A:Residues: 1-365 <BLA>
R:Mandi, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.
J. Mol. Biol. 257, 561-573, 1996
A:Title: Holiday junction resolvases encoded by homologous rsa genes in Escherichia
A:Reference number: S66579; MUID:96196428
A:Accession: S66594

A:Molecule type: DNA
A:Residues: 347-365 <MAH>
A:Cross-references: EMBL:X92587; NID:g1051136; PIDN:CAA6325.1; PID:g1051145
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.; Rose, D.D.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64787
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'NMIVAVTSGFFNNSSKGILT',1-325,'N',327-347,'EGANKSLI' <BLAT>
A:Cross-references: GB:AE000160; GB:U00096; NID:q1786751; PIDN:AAC73654.1; PID:g1786765
A:Experimental source: strain K-12, substrain M61655
C:Comment: In wild-type strains of *E. coli* K-12, the nmpc open reading frame is interrupted
in CS348, the ISS element has been deleted and nmpc is expressed.
C:Genetics:
A:Gene: nmpc
A:Map position: 12 min
A:Genome: cryptic lambdaoid prophage DLP12
C:Superfamily: outer membrane protein phor
C:Keywords: membrane protein; porin; trimer
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin nmpc #status predicted <MAT>

Query Match 11.88; Score 108; DB 1; Length 365;
Best Local Similarity 26.68; Pred. No. 0.064;
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;

```
QY      1 MKTKITLTV-AVSASSLLAMSANAAISY-----GNSADAOPVYGAK 39  
          |||::||::|||||::|||::|||:  
Db       1 MKKLTVAISAASVAALMSAQAIAETYNKDSNKLDLVGVNAKHFFSSNDADDGGTTYAR 60  
  
QY      40 IGVDAKDARKINKNRYAGIYACYNEDQNFGVAEAEVYSKAKEFNMGAVSFYKGDVASFAG 99  
          |||::|||::|::|::|::|::|::|::|::|:  
Db       61 LGFKGETIGINOLGLGFEGWME-YEFKGN--RAESOGSSKRDLRLFAAIK----FGDYG 111  
  
QY     100 TVRYAFNFTPEPYAKGL-----GIATKYKVY-----TSRN----- 130  
          :|::|::|::|::|::|::|::|::|::|:  
Db     112 SIDGRNNGVAYIDAGAWTDVLPERFGDTWTQTVDYEMTORATGYATTIRNDFPGLVDGINF 171  
  
QY     130 ATTYSNKSCKT---SLAGVGVEFRPLANNVGSEASYNTLSDDANAISLGAHLA 179  
          |::|::|::|::|::|::|::|::|::|:  
Db     172 AAOYQGKNDRSDFDNYTEGNGDGF-----GFSATYEY-----EGFGIGATYA 213
```

RESULT 9

S68072
major outer membrane protein P2, antigenic variant t1 - Haemophilus influenzae
C:/Species: Haemophilus influenzae
C:/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 08-Oct-1999
C:/Accession: S68072; S45437
R:Dultm, B.
submitted to the EMBL Data Library, June 1993
A:/Reference number: S68062
A:/Accession: S68072
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-371 <DUU>
A:/Cross-references: EMBL:X73386; NID:g510590; PIDN:CMA51803.1; PID:g510591
A:/Experimental source: isolate t1
A:/Mol. Microbiol. 11, 1181-1189, 1994
A:/Title: Antigenic cluster of non-encapsulated Haemophilus influenzae major outer membrane
A:/Reference number: S43699; MUID:94293786
A:/Accession: S45437
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 174-188;215-234;257-283 <DUW>
A:/Cross-references: EMBL:X73386
A:/Experimental source: isolate t1
A:/Keywords: membrane protein

	Query Match	11.3%	Score 103;	DB 2;	Length 371;	
	Best Local Similarity	24.8%;	Pred. No. 0.17;			
	Matches 53;	Conservative 29;	Mismatches 74;	Indels 58;	Gaps 12;	
QY	5	KTLLAVSASSLLAMSANMAISYGSNADAPPYGATIGOV-----DAKOING-----K	51			
		: : : : : : : : : : : : : : :				
Db	3	KTLLAIIIGAFAPASANAAYVYNNNEG-SKVELGRLSYIAEQSNNTYVDQKQHGALRNQ	61			
QY	52	NNAVYIVGYNDDQNF-----GYAEFYG----SDAEFPAAGVSPVAGDVKSPCAVCTYRYN	104			
		: : : : : : : : : : : : : : :				
Db	62	GSRFIKKITHNNGDFYKQGYLETRFISHYQDNADHPD-----DIT-----KYAVV	108			
QY	105	FLNTEPFYAKGLGIKATKY-DVTSRNATTYS--NKSDDTSLAGV-VGVGFPLANVGYEA	160			
		: : : : : : : : : : : : : :				
Db	109	TIGNKAFGVEVKIRAKTIAADITTSAEDEKREGYVLNNSKYLIRTMGNVTGYTFPKID--GLVL	166			
QY	161	SYNYLSEDA-----NAISIGA	176			
		: : : : : : : : : : : : : :				
Db	167	GANYLLAQRDTANPGKKGEVAAGSISNGVOYGA	200			

RESULT 10
 568069
 major outer membrane protein P2, antigenic variant t2 - Haemophilus influenzae
 C:Species: Haemophilus influenzae
 C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
 C:Accession: S68069; S43706
 R:Duim, B.
 submitted to the EMBL Data Library, June 1993
 A:Reference number: S68062
 A:Accession: S68069
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-371 <DU1>
 A:Cross-references: EMBL:X73383; NID:g475192; PIDN:CAA51800.1; PID:el192138; PID:g2655
 A:Experimental source: isolate t2
 R:Duim, B.; van Alphen, L.; Eljk, P.; Jansen, H.M.; Dankert, J.
 A:Title: Antigenic drift of non-encapsulated Haemophilus influenzae major outer membr
 A:Reference number: S43699; MUID:94293786
 A:Accession: S43706
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 174-188; 215-234; 257-283 <DUW>
 A:Cross-references: EMBL:X73383
 A:Experimental source: isolate t2
 C:Keywords: membrane protein

	Query Match	11.3%	Score 103;	DB 2;	Length 371;
	Best Local Similarity	24.8%;	Pred. No. 0.17;	Mismatches 74;	Indels 58; Gaps 12;
	Matches	53;	Conservative	29;	
OY	5 KTLVAVSASSLLAMSANNAISYGNSADAPPYGAKIGOV-----DAKQING----K 51	: : : : : : : : : : : :			
Dd	3 KTLALILYGAFAASANAAYVNNNG-SKEVLGGSLSYIAEGSNNTYDDQHQGHALNQ 61				
OY	52 NTATGYIYGINFDONE-----GYAEFFVC-----SDAKEFNAGVSPVKGDYKSFSGANGTYRYN 104	: : : : : : : : : : : :			
Dd	62 GSREFIKATHNNGDFYAGQYLETFRISHYDNDMDHFP-----DLTT-----KVAYV 108	: : : : : : : : : : : :			
OY	105 FINTPEFYANKGIGIAKTAV-DVTSNATTYS--NKSDKTSLAGG-VGVGFPIANVGVEA 160	: : : : : : : : : : : :			
Dd	109 TLGNNAFGEVYLGRAKTIADDTTSAEDDEKYLVANSKIIRTINGNVGTIFKGID--GLVL 166	: : : : : : : : : : : :			
OY	161 SYNVLSEDA-----NAISIGA 176	: : : : : : : : : : : :			
Dd	167 GANYLLAARDPTANPGKOGVEAAQSISNGYVGA 200	: : : : : : : : : : : :			

RESULT 11

A56152
major 25k outer membrane protein precursor - Brucella abortus
C:Species: Brucella abortus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 08-Oct-1999
C:Accession: A56152
R:de Wergifosse, P.; Lintermans, P.; Limet, J.N.; Cloeckaert, A.
J. Bacteriol. 177, 1911-1914, 1995
A:Title: Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton
A:Reference number: A56152; MUID:95204367
A:Accession: A56152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <DEA>
A:Cross-references: GB:X79284; NID:g769744; PIDN:CAA5872.1; PID:g769745

Query Match 11.2%; Score 102.5; DB 2; Length 213;
Best Local Similarity 27.1%; Pred. No. 0.1;
Matches 60; Conservative 22; Mismatches 60; Indels 79; Gaps 11;

Qy 1 MKTKTLAVSASSLLAMANAISYGSADA---QPYVGAKI----- 41
Db 1 MRLTKSLIVISA-ALPEPSATAF-----ADAIOEPVPAPAVEAPQYSWAGYTGLYL 54
Qy 41 --GGVDAK-----QINGKNTAVGIYAGYNFDON---FGVEAFVGSDAKEFNAGVSPYKG 90
Db 55 GYGNMKATSTVGSITKPDMDKAGAPAGNFFQDDQIVYGEVDAGYSMAKSKSGLEVKOG 114
Qy 91 DVKSFAGVGYRYNFINTPFYAKGKLG-----AKTKYDV 125
Db 115 FEGSLGARVGYDLNPV-MPYLTAGIAGSQIKLNGLDESKFRVGTAGAGLEAKLTQNI 173
Qy 126 TSR--NATYTSNKSDDKTSIAG-----GVGVGFK 151
Db 174 LGREYRYTYGKNK--YDLAGTYVRNKLDTQDIFVGIYK 212

RESULT 12
568062
major outer membrane protein P2 - Haemophilus influenzae (isolate d2 and isolate 3dr)
C:Species: Haemophilus influenzae
A:Variety: isolate d2; isolate 3dr
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
C:Accession: S68062; S68065; S43699
R:Duim, B.
submitted to the EMBL Data Library, June 1993
A:Reference number: S68062
A:Accession: S68062
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-386 <DUIM>
A:Cross-references: EMBL:X73376; NID:g475184; PIDN:CAA51796.1; PID:e1192124; PID:g265428
A:Experimental source: isolate d2
A:Accession: S68065
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-386 <DUIM>
A:Cross-references: EMBL:X73379; NID:g475184; PIDN:CAA51796.1; PID:e1192124; PID:g265428
A:Experimental source: isolate 3dr
R:Duim, B.; van Alphen, L.; Eljk, P.; Jansen, H.M.; Dankert, J.
Mol. Microbiol. 11, 1181-1189, 1994
A:Title: Antigenic drift of non-encapsulated Haemophilus influenzae major outer membrane
A:Reference number: S43699; MUID:94293786
A:Accession: S43699
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 89-105;135-153;178-196;223-236, 'GD', 239-246;269-275, 'HY', 278-298 <DUIM>
A:Cross-references: EMBL:X73376
A:Experimental source: isolate d2
C:Keywords: membrane protein

Query Match 11.1%; Score 101; DB 2; Length 386;

Best Local Similarity 24.1%; Pred. No. 0.26;
Matches 45; Conservative 28; Mismatches 74; Indels 40; Gaps 8;

Qy 5 KTLAVSASSLLAMANAISYGSADAQPYVGAKIGV-----DAKING-----K 51
Db 3 KTLAALIVGAFASANAAYVNNBG-TNVELGRLSTITEGNSGVDDQEQHGLRNA 61
Qy 52 NTAYGIYAGYNFDONFGE-----AEFVGSDAKEFNAGVSPYKGDKSFGAGTYRY 103
Db 62 GSRFHIAKTHNFGDGFYAGYLETRLVSDYPSSSDHFG-----GITTKYAY 108
Qy 104 NFINTPFYAKGLGIATKRV-DVTSRNATY---SNKSDKTSIAGVGVGFPLANV-GV 158
Db 109 VTLGNKAFGEVYKLRKATIDGITSADKEYGVLNKKIITLNGMTVGYTYKGIDGLDL 168
Qy 159 EASYNYL 165
Db 169 VLGANYL 175

RESULT 13
568063
major outer membrane protein P2, antigenic variant d3 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
C:Accession: S68063; S43700
R:Duim, B.
submitted to the EMBL Data Library, June 1993
A:Reference number: S68062
A:Accession: S68063
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-386 <DU3>
A:Cross-references: EMBL:X73377; NID:g475180; PIDN:CAA51794.1; PID:e1192122; PID:g265
A:Title: Antigenic drift of non-encapsulated Haemophilus influenzae major outer membr
A:Reference number: S43699; MUID:94293786
A:Accession: S43700
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 89-105;135-153;178-196;223-236, 'GD', 239-246;269-275, 'HY', 278-298 <DUIM>
A:Cross-references: EMBL:X73377
A:Experimental source: isolate d3
C:Keywords: membrane protein

Query Match 11.1%; Score 101; DB 2; Length 386;
Best Local Similarity 24.1%; Pred. No. 0.26;
Matches 45; Conservative 28; Mismatches 74; Indels 40; Gaps 8;

Qy 5 KTLAVSASSLLAMANAISYGSADAQPYVGAKIGV-----DAKING-----K 51
Db 3 KTLAALIVGAFASANAAYVNNBG-TNVELGRLSTITEGNSGVDDQEQHGLRNA 61
Qy 52 NTAYGIYAGYNFDONFGE-----AEFVGSDAKEFNAGVSPYKGDKSFGAGTYRY 103
Db 62 GSRFHIAKTHNFGDGFYAGYLETRLVSDYPSSSDHFG-----GITTKYAY 108
Qy 104 NFINTPFYAKGLGIATKRV-DVTSRNATY---SNKSDKTSIAGVGVGFPLANV-GV 158
Db 109 VTLGNKAFGEVYKLRKATIDGITSADKEYGVLNKKIITLNGMTVGYTYKGIDGLDL 168
Qy 159 EASYNYL 165
Db 169 VLGANYL 175

RESULT 14
568064
major outer membrane protein P2, antigenic variant d4 - Haemophilus influenzae
C:Species: Haemophilus influenzae

THIS PAGE BLANK (USPTO)

Accession	Length	Score	Expect	Ident	Accession	Length	Score	Expect	Ident
34	89	9.8	171	1	OMPX_ECOCI	100	10.0	171	1
					P36546	100	10.0	171	1
					escherichla	100	10.0	171	1

QY 59 AGYNFDQNFGEAEFVGSDAKEFN-AGVSPVKGDVKSFGAYGTYRYNEINTPTYAKGLG 1.17

Db 51 VGYAFDKNFR LAVDYTN YGKVTANYADVDVSLK GKSLGLTG FYD FDLADFKPYV--GVR 108

QY 118 IAKTKVDVTSRNATYTSNKSDDTSLAGV----GVGFKPLANGVEAS--YNTLSEDANA 171
DB 109 VSTNGADVTA-NARYYRIEAFATETRIGICALAGVQYKILDMVALNTIEYNRLASNVSD 167
QY 172 ISICAHIAF 180
DB 168 VGKAGLRP 176

RESULT 2

AIL_YEREN ID AIL_YEREN STANDARD; PRT; 178 AA.
AC P16454;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATTACHMENT INVASION LOCUS PROTEIN PRECURSOR.
GN AIL.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
RX MEDLINE; 90130261.
RA Miller V.L., Bilska J.B., Falkow S.;
RT "Nucleotide sequence of the Yersinia enterocolitica ail gene and
RT characterization of the Ail protein product."
RL J. Bacteriol. 172:1062-1069(1990).
CC -1- FUNCTION: THIS MEMBRANE-ASSOCIATED PROTEIN PROMOTES INVASION
CC OF PATHOGENIC BACTERIA INTO EUKARYOTIC CELLS BY AN UNKNOWN
CC MECHANISM.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOW FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
CC or send an email to license@isb-stb.ch).
CC -----
CC EMBL; M29945; AAA88694.1; -.
DR PIR; A35123; A35123.
DR PRINTS; PR00316; ENTEROYIROMP.
DR PROSITE; PS00694; ENT_VIR_OMP_1; 1.
DR PROSITE; PS00695; ENT_VIR_OMP_2; 1.
KW Outer membrane; Transmembrane; Signal; Virulence.
FT SIGNAL 1 23
FT CHAIN 24 178 ATTACHMENT INVASION LOCUS PROTEIN.
SQ SEQUENCE 178 AA; 19548 MW; D46FC92E2524F162 CRC64;

Query Match 13.5%; Score 123; DB 1; Length 178;
Best Local Similarity 29.6%; Pred. No. 0.0009;
Matches 55; Conservative 29; Mismatches 82; Indels 20; Gaps 9;

QY 5 KTLIAVSASSLLAMSAANAISYGSNADAPYVGAKIGQY--DAKOINGKNTAYGIAGYN 62
DB 3 KILL--ASSLLACLSIAVNVYMAESSISIGYASQSHVKENGTYLDDPKGNLKYRRE 59
QY 63 PDONGVEAEFVGS--DAKEFNAGVSPV-KGVKSFEGATGYRYNF-INTPFYAKGKLGIA 119
DB 60 LDDMGVIGSFAYTHGQYGFYFGSNKFGHDVYVS--TMGSPFINNEYVSLYGLLGA 117
QY 120 KTRVDVTSRNATY--SNKSDKTSIAGVGVGFKPLANGVEASINYISEDANAIS---L 174
DB 118 HGKV-----KASVFDESISASKTSMAAGVGFNPFPNVIDASYSKLDISIKVGTWML 172
QY 175 GAHLAF 180
DB 173 GAGYRP 178

RESULT 3
AIL_YERPS ID AIL_YERPS STANDARD; PRT; 182 AA.
AC 056957;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATTACHMENT INVASION LOCUS PROTEIN PRECURSOR.
GN AIL.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YPIII;
RX MEDLINE; 96294755.
RA Yang Y., Merriam J.J., Mueller J.P., Isberg R.R.;
RT "The psa locus is responsible for thermoinducible binding of Yersinia
RT pseudotuberculosis to cultured cells."
RL Infect. Immun. 64:2483-2489(1996).
CC -1- FUNCTION: THIS MEMBRANE-ASSOCIATED PROTEIN PROMOTES INVASION
CC OF PATHOGENIC BACTERIA INTO EUKARYOTIC CELLS BY AN UNKNOWN
CC MECHANISM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOW FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
CC or send an email to license@isb-stb.ch).
CC -----
CC EMBL; L49439; AAB3601.1; -.
DR PRINTS; PR00316; ENTEROYIROMP.
DR PROSITE; PS00694; ENT_VIR_OMP_1; 1.
DR PROSITE; PS00695; ENT_VIR_OMP_2; 1.
KW Outer membrane; Transmembrane; Signal; Virulence.
FT SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 182 ATTACHMENT INVASION LOCUS PROTEIN.
SQ SEQUENCE 182 AA; 20192 MW; EE7A4A20E4D8975C CRC64;

Query Match 12.7%; Score 115.5; DB 1; Length 182;
Best Local Similarity 27.8%; Pred. No. 0.0039;
Matches 52; Conservative 29; Mismatches 85; Indels 21; Gaps 7;

QY 5 KTLIAVSASSLLAMSAANAISYGSNADAPYVGAKIGQYDAKOINGKNTAYGIAGYNFD 64
DB 6 KILLVSSLACLSIAVNVYMAEBSISIGYASQSRV-KDGYKLDNPNPGFNLYKYEEN 64
QY 65 QNFGEAEFVGSDA-----KEFNAGVSPYKGDV-K-SFGATGYRYNF-INTPFYAKGKL 116
DB 65 NDMGV-----IGSFAQTRRGFEESVDGFKLIDGFKYYSYTAGVFR---INNEYVSLYGLL 117
QY 117 GIATKVDVTSRNATYTSNKSDDTSLAGVGVGFKPLANGVEASINYISEDANAIS--- 174
DB 118 GAGHGAKAKYS--IFGQBSRSKTSIAYAGLQFNPHNPFVIDASYSKLDVKKVTGM 175
QY 174 LGAHIAF 180
DB 176 LGAGYRP 182

RESULT 4
OM53_HAEIN ID OM53_HAEIN STANDARD; PRT; 359 AA.
AC P45996;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P5 PRECURSOR (OMP P5) (FIMBRIN).
GN OMPA OR OMP5.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-NH1 1128;
RX MEDLINE; 9422575.
RA Strakova T., Kolatukudy P.E., Murwin D., Billy J., Leake E.,
RA Lam D., Demaria T., Bakaletz L.;
RT "Role of fimbriae expressed by nontypeable Haemophilus influenzae in
RT pathogenesis of and protection against otitis media and relatedness
RT of the fimbria subunit to outer membrane protein A.";
RL Infect. Immun. 62:2002-2020(1994).
CC -1- FUNCTION: ACTS AS A FIMBRIAL SUBUNIT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; I08448; AAA24959.1; -.
DR PRINTS; PRO1021; OMPADOMAIN.
DR PROSITE; PS01068; OMPA; FALSE_NEG.
DR PRAM; PF01389; Ompa_membrane; 1.
KW Outer membrane; Transmembrane; Porin; Signal; Fimbria.
FT SIGNAL 1 21
FT CHAIN 22 359 OUTER MEMBRANE PROTEIN P5.
FT DISUFID 332 344 BY SIMILARITY.
FT DOMAIN 278 322 OMPA-LIKE.
SQ SEQUENCE 359 AA; 36340 MW; 576B1C59B4818C37 CRC64;

Query Match 12.4%; Score 113; DB 1; Length 359;
Best Local Similarity 28.2%; Pred. No. 0.013;
Matches 55; Conservative 18; Mismatches 80; Indels 42; Gaps 11;

QY 5 KTLAVASASSLLAMSANAISYGNADAPYVAKIGQ-----VDAKQING----- 51
DB 3 KTAIALVYAGLAASVAGAAQENTF-----YAGVKAQGGSHDGINNGALIKKGLSSSNY 58
QY 51 ---KNT-AYGIYAGYNF--DQNGEVEAEFVGS--AKENFAGVSPVKGADYKSFQAY--- 99
DB 59 GYRNFTFYGVFGYQJLQNDNFGIAELGYDDEGRALREKGRPKAKH--TNHGAYLSL 116
QY 99 -GTVRYNFTNPFYAKGKLGIAKTKVD--VTSRNATYISKSDKSLAG--VGVEFK 151
DB 117 KGYTE---VLDGLDYVYAGVALVSDYKEYEDANGTRDHRKGRHTARASGLFVAGAEYA 173
QY 152 PLANVGEASYNYS 166
DB 174 VLPFLAVRLLEYQWLT 188

RESULT 5
OM21_HAEIN
ID OM21_HAEIN STANDARD; PRT; 359 AA.
AC P43839;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2 OR HI0139.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

CC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PROTOBACTERIUM STRAIN S89 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32699; AAC21810.1; -.
DR TIGR; HI0139; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 359 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 359 AA; 39375 MW; 7E9EFFC2BD50314 CRC64;

Query Match 12.3%; Score 112.5; DB 1; Length 359;
Best Local Similarity 28.1%; Pred. No. 0.014;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

QY 5 KTLAVASASSLLAMSANAISYGNADAPYVAKIGOV-----DAKQING-----K 51
DB 3 KTAIALVYAGLAASVAGAAVYNNNG--TNVELGRLSLIAESNSVTVDNQOQHARLNQ 61
QY 52 NTAVGIYAGYEDQNF-----GVAEFV-----GSDAKFENAGVSPVKGADYKSFQAYGY 101
DB 62 GSRFIKATHNNGDGFYAGYLETFYTKASENGSD--NF-----GDITS-----KY 106
QY 102 RYNFINTPFYAKGKLGIAKTKVD--VTSRNATYIS--NKSDKTSLAG--VGVEFKPLANV 157
DB 107 AYVTLGNKAFGEVKGIRAKTIADGITSADKEYGVNLNSDYIPTSGNVYGYFKGIDGLV 166
QY 158 VEASY 162
DB 167 LGANT 171

RESULT 6
OM21_HAEIN
ID OM21_HAEIN STANDARD; PRT; 361 AA.
AC P20149;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 89079316.

```

RA Munson R.S. Jr., Tolan R.W. Jr.:
RT "Molecular cloning, expression, and primary sequence of outer
RL membrane protein P2 of Haemophilus influenzae type b.";
RN Infect. Immun. 57:88-94(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 90158127.
RA Munson R.S. Jr., Bailey C., Grass S.;
RT "Diversity of the outer membrane protein P2 gene from major clones of
RL Haemophilus influenzae type b.";
RN Mol. Microbiol. 3:1797-1803(1989).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 89173305.
RA Hansen E.J., Hasemann C., Clausell A., Capra J.D., Orth K.,
RT "Primary structure of the porin protein of Haemophilus influenzae
RT type b determined by nucleotide sequence analysis.";
RL Infect. Immun. 57:1100-1107(1989).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PHOTODACTERYUM STRAIN SS9 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03359; AAA24993.1; -.
DR EMBL; A09003; CA000819.1; -.
DR PIR; A30542; A30542.
DR PIR; S09622; S09622.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 361 OUTER MEMBRANE PROTEIN P2.
FT SEQUENCE 361 AA; 39701 MW; AF104C9CD8942D69 CRC64;
SQ
Query Match 12.3%; Score 112.5; DB 1; Length 361;
Best Local Similarity 28.1%; Pred. No. 0.014;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;
QY 5 KTLAVASASSLLMSANAISYGNADAPYVGAIGOV-----DAKOING-----K 51
DB 3 KTLALIVAFASAAANAAVYINNEG-TINVELGRLSIIIEOSNSTVYDNKKQHGALRNQ 61
QY 52 NPAVGIYAGYNFDNF---GVDAEFV-----GSDAKENAGVSPYKGVKSGFAGYGY 101
DB 62 GSRFHIKATFNFGDYAAGYLETRFVTKASENGSD--NF-----GDITS-----KY 106
QY 102 RYVFINTPFAKOKLGIATKVD-VTSRNAVTS--NKSPTSLAGG-VGVGKPLANVG 157
DB 107 AYTTLNKAKEGVEKLGRAKTIADGITSADKREXVGLNSDIPTISGNTVGYTFKGDGLV 166
QY 158 VEASY 162
DB 167 LGANY 171
RESULT 7
PORT BPPA2 STANDARD; PRT; 365 AA.
AC PORT BPPA2
ID P07238;
AC P07238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORIN PROTEIN LC PRECURSOR.
GN LC.
OS Bacteriophage PA-2.

```

```

OC viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86304457.
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and ompc outer membrane porin protein genes of
RL lambdaoid bacteriophage.";
RN J. Biol. Chem. 261:12723-12732(1986).
CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER
CC THAN 4000-6000 DALTONS).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
CC -1- MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
CC THE LYSOGENIC STATE. IN E. COLI THE EXPRESSION OF OMPC AND OMPF
CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHO FAMILIES OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02580; AAA32301.1; -.
DR PIR; D25647; MMBP2.
DR HSSP; P02931; IGFN.
DR PRINTS; PR00182; ECOLNIPORIN.
DR PRINTS; PR00183; ECOLIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN.
DR PFAM; PF00267; Gram-ve_porins; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23
FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN LC.
FT CONFLICT 99 99 H -> R (IN AA SEQUENCE).
FT SEQUENCE 365 AA; 40290 MW; 0FBC0531FB9C0205 CRC64;
SQ
Query Match 12.0%; Score 109; DB 1; Length 365;
Best Local Similarity 26.6%; Pred. No. 0.029;
Matches 62; Conservative 20; Mismatches 77; Indels 74; Gaps 11;
QY 1 MKTLTKILL-AVSASSLLMSANAISY-----GNSADAQPYVGAK 39
DB 1 MKRLTVAISAVASVLMAMSQAAEIYNKDSNKLIDYGVNAKHFFSSNDADGDTTYAR 60
QY 40 IGVNDAKOINGKMTAVGIYGVNFDFGVDAEFVGSDAKEFNAGVSPYKGVKSGFAGY 99
DB 61 LGFKGEQINDQLTGFGOWP-YEFKGN--RAESQSSKDKTHLARGLK-----FGDGI 111
QY 100 TYRYNFINTPFAKOKL-----GIATKVDY-----TSRN----- 130
DB 112 SIDYGRNYGVAVDIGAMTDVLPFGGDTWTQTVDYFMTGRTTGATYRRNDFGLVGLNF 171
QY 130 ATTYSNKSKDT---SLAGYGVGFKPLANYGVDASTNYLSSEDAANISLGAHLA 179
DB 172 AAQYOGKNDKSDPDNYTEGNGDGF-----GFSATVEY-----EGFGIGATYA 213
RESULT 8
PAGC_SALTY STANDARD; PRT; 188 AA.
AC PAGC_SALTY
ID PAGC_SALTY
AC P23988;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VIRULENCE MEMBRANE PROTEIN PAGC PRECURSOR.
GN PAGC.

```

```

OS Salmonella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae:
OC Salmonella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 14028;
RX MEDLINE: 91100323.
RA Pulkkinen W.S., Miller S.I.;
RT "A Salmonella typhimurium virulence protein is similar to a Yersinia
RT enterocolitica invasion protein and a bacteriophage lambda outer
RT membrane protein."
RL J. Bacteriol. 173:86-93(1991).
CC -1- FUNCTION: ESSENTIAL FOR FULL VIRULENCE AND SURVIVAL WITHIN
CC MACROPHAGES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE ATL/OMPX/PAGC/LOM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55546; AAA27179.1; -.
DR PIR: A39185; A39185.
DR STYGENE: SG77777; PAGC.
DR PRINTS: PR00316; ENTEROVIROMP.
DR PROSITE: PS00694; ENT_VIR_OMP_1; 1.
DR PROSITE: PS00695; ENT_VIR_OMP_2; 1.
KW Outer membrane; Transmembrane; Signal; Virulence.
FT SIGNAL 1 23
FT CHAIN 1 188
FT SEQUENCE 188 AA; 20574 MW; B23826FA2B62DDEF CRC64;
SQ
Query Match 11.88; Score 108; DB 1; Length 188;
Best Local Similarity 24.28; Pred. No. 0.017;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;
QY 1 MKLTLLAVASSILLASANAISYGSADAQPYGAKIGQYAKOINGRKTAAGIAG 60
DB 1 MKNIIILSTLVTTVNVNAQADNAPSVGIARYAQSKV---QDEKNIKGVAVK----- 54
QY 61 YNFDNFEVGEAFVGSDAKEFNAGVSPYKGVKSEGA-----YGT----- 101
DB 54 -----RYEDDSVYSFISLSIYLXGDRQASGVPEPGIHHDKFEYKYGSLMVG 102
QY 101 -YRNFTTPPYANGKIGIAK-TKVDVTSRNATYTSNK--SDKTSLAGGVGVKPLANV 156
DB 103 AYRSL-DNFSILYALAGVGTAKTEKSHSTQDGSFNKISSRKITGFAGAGVQMPLENI 161
QY 157 GVEASY---NYLSEDAANISGAHLAF 180
DB 162 VVDVGEGSNISTKINGENVGVGYRF 188

```

```

RX MEDLINE: 86304457.
RA Blasband A.J., Maricotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and mmpc outer membrane porin protein genes of
RT lambdaoid bacteriophage."
RL J. Biol. Chem. 261:12723-12732(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federpspiel N., Hyman R., Kalman S., Komp C., Kirdi O., Lew H., Lin D.,
RA Nannath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-17 FROM N.A.
RC STRAIN-JL5502;
RX MEDLINE: 94335635.
RA Coll J.L., Heyde M., Portaller R.;
RT "Expression of the mmpc gene of Escherichia coli K-12 is modulated by
RT external pH. Identification of cis-acting regulatory sequences
RT involved in this regulation."
RL Mol. Microbiol. 12:83-93(1994).
RN [5]
RP SEQUENCE OF 347-365 FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 96196428.
RA Mandi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;
RT "Holiday junction resolvases encoded by homologous rusa genes in
RT Escherichia coli K-12 and phage 82."
RL J. Mol. Biol. 257:561-573(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- MISCELLANEOUS: IN WILD-TYPE STRAINS OF E. COLI K12, THE NMPC OPEN
CC READING FRAME IS INTERRUPTED BY AN IS5 INSERTION AND GENERATES A
CC HYBRID OPEN READING FRAME THAT IS NOT EXPRESSED. HOWEVER, IN
CC MUTANT STRAIN CS348, THE IS5 ELEMENT HAS BEEN DELETED AND NMPC IS
CC EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE OMPX/PHO FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annotate/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13457; AAA23728.1; ALT_SEQ.
DR EMBL: AE000160; AAC73654.1; ALT_SEQ.
DR EMBL: U82598; AAB40749.1; ALT_SEQ.
DR EMBL: Z35442; CAAB4594.1; -.
DR EMBL: X92587; CAAB3325.1; -.
DR PIR: A25647; MMECNC.
DR HSSP: P02931; IGFN.
DR ECOGENE: EG10659; NMPC.
DR PRINTS: PR00182; ECOLIPORIN.
DR PRINTS: PR00183; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
DR PFAM: PF00267; Gram-ve_Porins; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23
FT CHAIN 1 365
FT CONFLICT 326 326 K -> N (IN REF. 2 AND 3).
FT SEQUENCE 365 AA; 40316 MW; 6E512BD4647B4F8 CRC64;

```

```

Query Match      11.8%; Score 108; DB 1; Length 365;
Best Local Similarity 26.6%; Pred. No. 0.035;
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;

OY 1 MKTLAKTL-AVSASLLAMSANAAISSY-----GNSADAQPVYGA 39
DB 1 MKLTVAISNAVASVMAASQAELIYNDSNKLDIKGVNAKHRYSSNDADGDTTYAR 60
OY 40 ISQVDAKQINGKNTAYGIAGYVFNQDFGEAEFVSGDAKEFNAGVSPYKGDVKSFGAYG 99
DB 61 LQPKGRTQINDQLTGFGQWE-YEFKGN---RAESQSSKDKTRLAPAGLK-----FGDYG 111
OY 100 TYRYNININPFPAKGL-----GIATKYDV-----TSRN----- 130
DB 112 SIDYGRNYGVAYDIDIGAMTDVLPFGGDTWTQTDVFMTORATGVATYRNNDPFGLDVGLNF 171
OY 130 ATTYSKSPDKT---SLAGVGVGFKPLANVGVEASYNYLSEDANAISLGHILA 179
DB 172 MAOYGKNDRSPEDNTEBNGDGF-----GFSATYR-----EGFGIGATYA 213

RESULT 10
OM31_BRUME STANDARD; PRT: 240 AA.
AC 045322;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 31 KD OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP31.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-16M.
RX MEDLINE; 96355886.
RA Vizcaino N., Cloeckaert A., Zygmunt M.S., Dubray G.;
RT "Cloning, nucleotide sequence, and expression of the Brucella
RT melitensis omp31 gene coding for an immunogenic major outer membrane
RT protein.";
RT Infect. Immun. 64:3744-3751(1996).
CC -1- FUNCTION: MAJOR OUTER MEMBRANE PROTEIN ASSOCIATED WITH
CC -1- PEPTIDOGLYCANS. MAY FUNCTION AS A PORIN.
CC -1- SUBUNIT: OLIGOMERIC.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF076290; AAB3693.1; -.
DR PFM; PF01389; OmpA membrane; 1.
KW Antigen; Outer membrane; Porin; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 240 31 KD OUTER-MEMBRANE IMMUNOGENIC PROTEIN.
FT DOMAIN 48 83 EPITOPE RECOGNIZED BY THE MONOCLONAL
FT ANTI-BODY A59/10F09/G10.
SQ SEQUENCE 240 AA; 25323 MW; 21C6E5EC479F66A5 CRC64;

Query Match      11.5%; Score 105; DB 1; Length 240;
Best Local Similarity 24.4%; Pred. No. 0.04;
Matches 51; Conservative 29; Mismatches 85; Indels 44; Gaps 9;

OY 4 KTLTLLAVSASSLLAMSANA---ISYGNSADAQ-----YVAKIG----- 42
DB 1 MKSVLLASIAAMFATISAMADADVVSSEPSAPTAAPVDTFSWGTGYIGINAGYAGKFKHP 60

```

```

OY 42 ----QVDAKQINCK-----NTAYGIAGYVFNQDFV----EAEFVSGDAK-EFNAGVS 86
DB 61 SRFDKEDNEQVSGSLDVTAGGFGVQAGYNMQLDNGVYLGAETDPQGSVTSISAGAS 120
OY 87 PVKGDVKS-FGAGTYRYNFINP-----FYAKGKIGIAKTKYDV-SRNATTYSKSKDK 139
DB 121 GLEKAETEVEMWGYRARGYATERLWYGTGGLAYGKVSAFNIGDDASALHTWSDK 180
OY 140 TSLAGVGVGFKPLANVGVEASYNYLSED 168
DB 181 TRAGMTLGAGAEYAINNMNTLSEYLYTD 209

RESULT 11
OM2B_HAEIN STANDARD; PRT: 363 AA.
AC 048220;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3230B.
RX MEDLINE; 94018553.
RA Duim B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae.";
RT Microb. Pathog. 14:451-462(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X7391; CAA51808.1; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 363 OUTER MEMBRANE PROTEIN P2.
FT SEQUENCE 363 AA; 39873 MW; 2739FDBF42A3051 CRC64;

Query Match      11.5%; Score 105; DB 1; Length 363;
Best Local Similarity 26.9%; Pred. No. 0.062;
Matches 52; Conservative 29; Mismatches 78; Indels 34; Gaps 11;

OY 5 KTLTLLAVSASSLLAMSANAISYGNSADAQPVYGAKIGOV-----DAKQING-----K 51
DB 3 KTLAALIVGAFASANAAVVYNNEG-TNVELGGRSLIAEGSNSTIKQKQGHGALRMQ 61
OY 52 NTAYGIAGYVFNQDF-----GVAEFVSGDAKEFNAGVSPYKGDVKSFGAYGI-YRYNPI 106
DB 62 SSRFHAKATHNFGDGFYAGGYLETRLVSAQ-----SGTESD--NFGHIIIKRYAVYVL 111
OY 107 NPFYVKGKGIGIAKTKYVD-VTSRNATTYS--NKSDKTSLAGG-VGVGFKPLANVGVEASY 162
DB 112 GKRAFEVKLGRAKTTADGITSAEDKEGYLVNSKIYPTNGNTGTYTTERGIDGLVIGANT 171
OY 163 NYLSEDANAISLG 175
DB 172 -LLAQERNKRYGTG 183

RESULT 12
OM25_HAEIN

```

```

ID OM25_HAEIN STANDARD; PRT; 371 AA.
AC P46027;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE: 94293786.
RA Duim B., van Alphen L., Bijl P., Jansen H.M., Dankert J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein P2 in patients with chronic bronchitis is
RT caused by point mutations."
RL MOL. Microbiol. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE: 94018553.
RA Duim B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae."
RL Microb. Pathog. 14:451-462(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMP1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73386; CAAS1803.1; -.
DR EMBL: X73383; CAAS1800.1; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT CHAIN 1 20
FT SIGNAL 21 371
FT VARIANT 184 184 K -> Q (IN T2).
FT VARIANT 224 224 A -> T (IN T2).
FT VARIANT 273 273 S -> T (IN T2).
SQ SEQUENCE 371 AA; 40901 MW; D0F252D7A6E5CDA6 CRC64;

Query Match 11.3%; Score 103; DB 1; Length 371;
Best Local Similarity 24.8%; Pred. No. 0.093;
Matches 53; Conservative 29; Mismatches 74; Indels 58; Gaps 12;

QY 5 KTLIAVSASSLLMSANAISYGSNADQPYVAKIGOV-----DAKQING-----K 51
DB 3 KTLAALIVGFAASAAVAAYVANNESG-SKVELGRLSVIAEQSNNTVDDQOQOGALRNQ 61
QY 52 NTAVGIYAGYNFDPNFGVGEAEFVG---SDAKFEFNAGVSPYKGDVSKFGAGTYRYN 104
DB 62 GSRHIAKATINFGGFGFYAGQYLETFRISHYQDNADHFD-----DITF-----KRAY 108
QY 105 FINTPFYAKKGLIAKTKV-DVTSRNATYS--NKSOKTSLIAG-VGVGKPLANVGVEA 160
DB 109 TLGKAKAGEVKLGRAKTIADITSAEDKEVGLNNSKYIRTNMTVGTYTGID--GLVL 166
QY 161 SYNTLSEDA-----NAISLGA 176
DB 167 GANYLLAQARDTANPGKKEGVAAGSISNGVQVA 200

```

```

ID OM25_BRUBAB STANDARD; PRT; 213 AA.
AC Q44664;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KD OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-544S;
RX MEDLINE: 95204367.
RA de Wergifosse P., Lintermans P., Lmet J.N., Cloeckaert A.;
RT "Cloning and nucleotide sequence of the gene coding for the major 25-
RT kilodalton outer membrane protein of Brucella abortus."
RL J. Bacteriol. 177:1911-1914(1995).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79284; CAAS5872.1; -.
DR PFAM: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 213
SQ SEQUENCE 213 AA; 23052 MW; 2328515F1F794BC7 CRC64;

Query Match 11.2%; Score 102.5; DB 1; Length 213;
Best Local Similarity 27.1%; Pred. No. 0.057;
Matches 60; Conservative 22; Mismatches 60; Indels 79; Gaps 11;

QY 1 MKTLKTLIAVSASSLLMSANAISYGSNADQPYVAKI----- 41
DB 1 MKTLKSLIYISA-ADLPESATAF-----ADAIDQDPVPAPVEYAPQYSNAGTYTGLYL 54
QY 41 --GOVDN-----QINGKNTAYGIYAGYNFDPN--FGVEAEFVGSDAKEFNAGVSPYK 90
DB 55 GYGNKKAKTSTVSGSIKPDMDKAGAFAGNPFQDDIYGVGEDAGYSMAKKSKDGLVKG 114
QY 91 DVKSEFAGYTRYNFINTPFYAKKGLI-----AKTKVDV 125
DB 115 FEGSLGKAVGYDLDPV-MPYLTLAGIAGQIKRLNGLNGLDESKFRVGTAGLEAKLTNDI 173
QY 126 TSR---NATYYSNKSOKTSLIAG-----GVGVGFK 151
DB 174 LGRVEYRYTQIGKN--YDLAGTTVRNKLDTPQDIRVIGIKY 212

```

RESULT 13
OM25_BRUBAB

```
RX MEDLINE; 9535366.
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummur M.K., Blaser M.J.,
RT Cover T.L.;
RT "Mosaicism in vaccinating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vacA types with cytotoxin production and
RT peptic ulceration."
RL J. Biol. Chem. 270:17771-17777(1995).
CC -1- FUNCTION: INDUCES VACCINATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL; U29401; AAA86834.1; -.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 2 VACUOLATING CYTOTOXIN.
FT PROPEP ? 1310 POTENTIAL.
SQ SEQUENCE 1310 AA; 14198 MW; 1BC21FE3D435F981 CRC64;

Query Match 11.2%; Score 102.5; DB 1; Length 1310;
Best Local Similarity 22.7%; Pred. No. 0.4; Mismatches 74; Indels 53; Gaps 10;
Matches 48; Conservative 36;

OY 9 AVSASSLLAMSAANAISYGNADAPYVGAKI-----GQVDAKQINGKN 52
DB 1049 AIGTS-LNNGSNASL-YGTSAVDAYLNGEYEAIVGFGSYGSFSGNSQANSLSNGANN 1106
OY 53 TAYGIYAGNFDQN-FGVAF-FVGSDAKEFNAGVSPVKDYKSF-----GAYGT 100
DB 1107 TNEGYSTRIFAQHEDEFAQALGSDSLNFKSALLDLNQSYYHLYATSATTRASYG- 1166
OY 101 YVNTINPFYKAGLGIKTKYVDTSRNATYTSNKSDDKTSLAGVGVGFPLANVGVEA 160
DB 1166 YVAFPRNALVLPKPSGVSYNHL-----GSTNFKSNSNOVALSSSSOHLNANAVEA 1220
OY 161 SYNY-----LSEDA-----MAISL 174
DB 1221 RXYGDTSYFYMNAGVLOEFARFGSNNNAVSL 1251

RESULT 15
OM24_HAEIN STANDARD: PRT; 386 AA.
AC P46026;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D2, D3, D3R, AND D4;
RX MEDLINE; 94293786.
RA Duim B., van Alphen L., Eljk P., Jansen H.M., Dankert J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein P2 in patients with chronic bronchitis is
RT caused by point mutations."
RL Mol. Microbiol. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D2, D3, D3R, AND D4;
RX MEDLINE; 94018553.
```

```
RA Duim B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae."
RL Microb. Pathog. 14:451-462(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- MISCELLANEOUS: THE SEQUENCE OF STRAIN D2/D3R IS SHOWN HERE.
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN S59 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL; X73376; CAAS1793.1; -.
DR EMBL; X73377; CAAS1794.1; -.
DR EMBL; X73378; CAAS1795.1; -.
DR EMBL; X73379; CAAS1796.1; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 386 OUTER MEMBRANE PROTEIN P2.
FT VARIANT 284 284 D -> G (IN D3).
FT VARIANT 284 285 DD -> GG (IN D4).
SQ SEQUENCE 386 AA; 41905 MW; 05F8909A2382A1F3 CRC64;
```

```
Query Match 11.1%; Score 101; DB 1; Length 386;
Best Local Similarity 24.1%; Pred. No. 0.14; Mismatches 74; Indels 40; Gaps 8;
Matches 45; Conservative 28;

OY 5 KTLAVASASSLLAMSAANAISYGNADAPYVGAKIGOV-----DAKING-----K 51
DB 3 KTLALIVGAFAPASANAAYVYNNEG-TNVELGGRSLITTEOSNSTVDDQDQHGALRNA 61
OY 52 NTAYGIYAGNFDQNGVE-----AEFVGSDAKEFNAGVSPVKDYKSGAGYTRY 103
DB 62 GSRFHIKATHNFGDGFYAGGYLETRLVSDYPGSSSDHFG-----GITKKYV 108
OY 104 NFINTPFYKAGLGIKTKYVD-VTSRNATY---SKSKTSLSAGVGVGFPLANV-GV 158
DB 109 VTLGNKAFGEVRLGKRAKTIADGITSADREYGLNKKYILTNQNTVGYTKGIDLDGL 168
OY 159 EASYNYL 165
DB 169 VLGANYL 175
```

Search completed: May 20, 2000, 12:22:46
Job time: 3960 sec

THIS PAGE BLANK (USPTO)


```

RESULT 2
ID P96774 PRELIMINARY; PRT; 407 AA.
AC P96774;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, last annotation update)
GN OMPA2.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3500;
RX MEDLINE; 97197543.
RA KLESNEY-TAIT J., HILTKE T.J., MACIVER I., SPINOLA S.M., RADOLF J.D.,
RA HANSEN E.J.;
RT "The major outer membrane protein of Haemophilus ducreyi consists of
RT two ompA homologs.";
RL J. Bacteriol. 179:1764-1773(1997).
DR EMBL; U60646; AAB49274.1; -.
DR HSSP; P02934; 1BXW.
DR PFAM; PF00691; OmpA.1.
DR PFAM; PF01389; OmpA_membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
SQ SEQUENCE 407 AA; 44702 MW; 4C2BA5D7 CRC32;

```

```

Query Match 14.1%; Score 129; DB 2; Length 407;
Best Local Similarity 27.6%; Pred. No. 0.0056;
Matches 64; Conservative 27; Mismatches 67; Indels 74; Gaps 13;

```

```

QY 2 MKTLKTLAVSASSLLAMSAANAISYGNSDADQPYGAKIG-----QVDAKQINGN 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 KLVYLAIVASAR-VATAAPQADTF-----YGAKKAGMSFHHGICQFQDKYNNRKH 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 53 TA-----YGIYAGYN-FDGN-FGVEAEFVSD-----AK-----EENAGV 85
   || || || || || || || || || || || || || || || || || || || ||
DB 54 TAYMEGKLINIRDSVYGVFAGYQIFNQNKAGLAELGVDYVGQLSAKFKHAENKNSIQ 112
   || || || || || || || || || || || || || || || || || || || ||
QY 86 SPVKGDKVRFSGAGYGYR-----NFINTP-----FYAKGKIGIAKTKVDYS 127
   || || || || || || || || || || || || || || || || || || || ||
DB 113 FPYVSAKMKRADNDNEEYVVRHTAGVNSLSKSYELLDPDLDAVAKVGIGLARNDYSVP 172
   || || || || || || || || || || || || || || || || || || || ||
QY 128 RNATYSNKSDDT-----SLAGVGVGFKPLANVCVEASYNVLSDDAN 170
   || || || || || || || || || || || || || || || || || || || ||
DB 173 KKKAV-NNKTEVKNNRYDVKPSLLIGAGFEYAIIPELALRYEYQILNRYGN 223
   || || || || || || || || || || || || || || || || || || || ||

```

```

RESULT 3
ID Q44298 PRELIMINARY; PRT; 339 AA.
AC Q44298;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMPA1.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC Aeromonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIMB 1102;
RX MEDLINE; 96198165.
RA COSTELLO G.M., VIPOND R., MACINTYRE S.;
RA "Aeromonas salmonicida possesses two genes encoding homologs of the
RT major outer membrane protein, OmpA.";
RL J. Bacteriol. 178:1623-1630(1996).
DR EMBL; X91983; CAA63036.1; -.
DR PFAM; PF00691; OmpA.1.
DR PFAM; PF01389; OmpA_membrane; 1.

```

```

DR PRINTS; PR01021; OMPADOMAIN.
KW Signal.
FT SIGNAL.
FT CHAIN
SQ SEQUENCE 339 AA; 35807 MW; B3F36814 CRC32;

```

```

Query Match 13.4%; Score 122.5; DB 2; Length 339;
Best Local Similarity 24.1%; Pred. No. 0.015;
Matches 46; Conservative 35; Mismatches 57; Indels 53; Gaps 8;

```

```

QY 1 MKTLKTLAVSASSLLAMSAANA--ISYGNSDADQPY-----VGAKIQVDAKQINGN 52
   || || || || || || || || || || || || || || || || || || || ||
DB 2 MKWAPSLAIAMAAAGCATAAHADIDYFCAGAGAAHFNGINIKIEGVKNGEENAAAANA-- 60
   || || || || || || || || || || || || || || || || || || || ||
QY 53 TAYGIYAGYFQNGVEAEFV-----GSDAKEF-NAGVS-----PYKGVKSGAGY 99
   || || || || || || || || || || || || || || || || || || || ||
DB 60 -----FVGTFNEDNFGSELGYLTGRGNTDGNRYENQGATLSGIARLPUGDPSAFABEG 114
   || || || || || || || || || || || || || || || || || || || ||
QY 100 TYRYNFINTPFYAKGKIGIAKTKVDVTSRNATYYSNKSDDKTSLAGVGVGFKPLANVGE 159
   || || || || || || || || || || || || || || || || || || || ||
DB 115 AY-----WAHTDGMGTSDFRKV-----SPIAGLGYTYQVNDALDIQ 149
   || || || || || || || || || || || || || || || || || || || ||
QY 160 ASYNYLSDDAN 170
   || || || || || || || || || || || || || || || || || || || ||
DB 150 ARRYRMWDVAD 160
   || || || || || || || || || || || || || || || || || || || ||

```

```

RESULT 4
ID Q54470 PRELIMINARY; PRT; 172 AA.
AC Q54470;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE OUTER MEMBRANE PROTEIN PRECURSOR.
GN Omp4.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-N28B;
RX MEDLINE; 96036211.
RA GUASCH J.F., FERRER S., ENFEDAQUE J., VIREJO M.B., REGUE M.;
RA "A 17 kDa outer-membrane protein (Omp4) from Serratia marcescens
RT confers partial resistance to bacteriocin 28b when expressed in
RT Escherichia coli.";
RL Microbiology 141:2535-2542(1995).
DR EMBL; Z37157; CAA85513.1; -.
DR PFAM; PF01389; OmpA_membrane; 1.
DR PRINTS; PR00316; ENTEROVIROMP.
KW Signal; Outer membrane.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 172 AA; 18434 MW; 7C38C7F8 CRC32;

```

```

Query Match 12.9%; Score 118; DB 2; Length 172;
Best Local Similarity 25.0%; Pred. No. 0.014;
Matches 48; Conservative 26; Mismatches 86; Indels 32; Gaps 7;

```

```

QY 1 MKTLKTLAVSASSLLAMSAANAISYGNSDADQPYGAKIGQVDAKQINGKRNAGIYAG 60
   || || || || || || || || || || || || || || || || || || || ||
DB 1 MKKIACLSAVAACVY-----AVSAGITVPAQOSTVYSAGYAGQDGGVANKKADGFNLKTR 53
   || || || || || || || || || || || || || || || || || || || ||
QY 61 YNFQDN-FGVEAEFVGSDAKEFNAGVSPYKGVKSGFAGYTYRYNFINT-PFY----- 112
   || || || || || || || || || || || || || || || || || || || ||
DB 54 YEDNNPLGVISF-----THLEKDGSDGFEYNNKQAYNSISAGPAYRINDMAS 101
   || || || || || || || || || || || || || || || || || || || ||
QY 112 AKGKIGIAKTKVDVTSRNATYYSNKSDDKTSLAGVGVGFKPLANVCVEASY---NYLSDD 168
   || || || || || || || || || || || || || || || || || || || ||
DB 102 IYGLVGLGFGKFTTNAQNGTSSRHDTAD-YGFTYAGAGIQLPFIENVALDVGYEQNRIRSYD 160
   || || || || || || || || || || || || || || || || || || || ||

```

QY 169 ANAISGAHLAF 180
Db 161 VGTWNGVGYRF 172

RESULT 5

085182 PRELIMINARY; PRT; 572 AA.
AC 085182;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DE FLAGELLIN A.
GN FLAA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
RN Campylobacter.
RP [1]
RC SEQUENCE FROM N.A.
RC STRAIN-D2290;
RA MEINERSMANN R.J., HIETT K.L.;
RT "Concerted evolution of duplicate fla genes in Campylobacter."
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF050188; AAC25641.1;
DR PFAM: PF00700; Flagellin_C; 1.
DR PFAM: PF00659; Flagellin_N; 1.
DR PRINTS: PR00207; FLAGELLIN.
SQ SEQUENCE 572 AA; 59021 MW; 3A8B8A00 CRC32;

Query Match 12.8%; Score 117; DB 2; Length 572;

Best Local Similarity 24.3%; Pred. No. 0.079; Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AATSYGNSADQAPVYAKTIGVDKQ-----INGKMTAVGIYA----- 60
Db 242 AAVAGATSDTFAINGVYKGVYDKDGDANGALVAALNSVYKDTTGVESIDANGQLLITS 301
QY 60 ---GYNFDQNGVEAEFVGSDAKEFNAGVSPVKGDK-----SFGAYGYRY 103
Db 302 REGGIRKIDGNIIGGA-FINADMKENYGRSLVKNKGDILISSNLSAGFGA----- 355
QY 104 NFINTPYAKGLGIATK--VDYTSRNATTYSNKSDDKTSIAG-----GVGVGFKP 152
Db 355 ---TOFISQASVSLRESKQIDANIDAMGFGSANKGVVLGYSSVAYMSAGSGFSS 410
QY 153 LANGVEASVNYLSEDAANAISLGA 176
Db 411 GSGYVSGSKNYSFGFANAIAISA 434

RESULT 6

030689 PRELIMINARY; PRT; 572 AA.
AC 030689;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DE FLAGELLIN A.
GN FLAA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
RN Campylobacter.
RP [1]
RC SEQUENCE FROM N.A.
RC STRAIN-D1118;
RA MEINERSMANN R.J., HIETT K.L.;
RT "Concerted evolution of duplicate fla genes in Campylobacter."
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 101-169 FROM N.A.
RC STRAIN-D224;

RA MEINERSMANN R.J., HIETT K.L., FIELDS P.I., HIETT K.L.;
RL J. Clin. Microbiol. 0:0-0(1997).
DR EMBL: AF050197; AAC25650.1;
DR EMBL: AF015092; AAB69354.1;
SQ SEQUENCE 572 AA; 59049 MW; A624D1EE CRC32;

Query Match 12.8%; Score 117; DB 2; Length 572;

Best Local Similarity 24.3%; Pred. No. 0.079; Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AATSYGNSADQAPVYAKTIGVDKQ-----INGKMTAVGIYA----- 60
Db 242 AAVAGATSDTFAINGVYKGVYDKDGDANGALVAALNSVYKDTTGVESIDANGQLLITS 301
QY 60 ---GYNFDQNGVEAEFVGSDAKEFNAGVSPVKGDK-----SFGAYGYRY 103
Db 302 REGGIRKIDGNIIGGA-FINADMKENYGRSLVKNKGDILISSNLSAGFGA----- 355
QY 104 NFINTPYAKGLGIATK--VDYTSRNATTYSNKSDDKTSIAG-----GVGVGFKP 152
Db 355 ---TOFISQASVSLRESKQIDANIDAMGFGSANKGVVLGYSSVAYMSAGSGFSS 410
QY 153 LANGVEASVNYLSEDAANAISLGA 176
Db 411 GSGYVSGSKNYSFGFANAIAISA 434

RESULT 7

051841 PRELIMINARY; PRT; 346 AA.
AC 051841;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DE OUTER MEMBRANE PROTEIN 34 PRECURSOR.
GN OMP34.
OS Actinobacillus actinomycetemcomitans
OC (Haemophilus actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCTC 9710;
RX MEDLINE: 98084499.
RA WHITE P.A., NAIR S.P., KIM M.J., WILSON M., HENDERSON B.;
RT "Molecular characterization of an outer membrane protein of
RT Actinobacillus actinomycetemcomitans belonging to the OmpA family."
RL Infect. Immun. 66:369-372(1998).
DR EMBL: AF005079; AAC00068.1;
DR HSSP: P02934; 1BXW.
DR PROSITE: PS01068; OMPA; 1.
DR PFAM: PF00691; OmpA; 1.
DR PFAM: PF01389; OmpA-membrane; 1.
DR PRINTS: PR01021; OMPADOMAIN.
KW SIGNAL.
FT SIGNAL. 1
SQ SEQUENCE 346 AA; 36905 MW; 8A80E0A3 CRC32;

Query Match 12.7%; Score 116; DB 2; Length 346;

Best Local Similarity 25.3%; Pred. No. 0.05; Matches 49; Conservative 22; Mismatches 75; Indels 48; Gaps 9;

QY 5 KTLAVSASSILLAMSANAISYGNADQAPVYAKTIGVDA-----KQ-----INGKNT 53
Db 3 KTAIALATAGLAATAVQAQAPANTF---YAGAKAGASSHHGUNGKQGVSTINRNS 58
QY 54 AYGIAGYNFQNGVEA--EFVGSDAKEFNAGVSPVKGDKSFGAYGYRYNFINTPPY 111
Db 59 AYGVFGYQIDNFAVEGYEYFGRSKAKVN-----GAQ-RFRHTAIGTTLA 104
QY 112 AK-----GKIGIAKTKVDYTSRNAT---TYSNKSDDKTSIAGGVGVGFKPLANV 156

Db 105 LKASPVLDNLVYGVGALLRSDYKVGQANKPDRYHMLKVPVAGGEVAILP--EL 162
 QY 157 GVEASNYLSEDA 170
 Db 163 ALRAEYQWVSRYGN 176

RESULT 8

Q46679 PRELIMINARY: PRT: 250 AA.

AC 046679: PRELIMINARY: PRT: 250 AA.
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE HEAT RESISTANT AGGLUTININ 1.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O9:H10:K99;
 RX MEDLINE: 95012721.
 RA LUTWYCHE P., RUPPS R., CAVANAGH J., WARREN R.A., BROOKS D.E.;
 RT "Cloning, sequencing, and viscometric adhesion analysis of heat-
 resistant agglutinin 1, an integral membrane hemagglutinin from
 RT Escherichia coli O9:H10:K99."
 RL Infect. Immun. 62:5020-5026(1994).
 DR EMBL: U07174; AAC13752.1; -;
 DR PFAM: PF01389; OmpA_membrane; 1.
 FT CHAIN 26
 SQ SEQUENCE 250 AA; 27477 MW; 9673F116 CRC32;
 HEAT RESISTANT AGGLUTININ 1.

Query Match 12.0%; Score 109.5; DB 2; Length 250;
 Best Local Similarity 21.0%; Pred. No. 0.11; Mismatches 75; Indels 71; Gaps 10;

Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;
 QY 1 MKTLKTLAVSASSLAMSANAIS-----XGNSADAQP 34
 Db 1 MEMKNVIAVSLAMAGMSTQALADESKTGFYVGKAGASVSLADQRFSLNGEETSK 60
 QY 35 YVAKTIGVDAKOINGKNAY--GIYAGYNPDONGV---EAEPV--GSDAKEFNAGVS 86
 Db 61 YKGG-----DGHDFVFSGIAAGYDFYPOFSIPVTELEFYARGLKADSKYNVDKD 110
 QY 87 PVKG-----DVKS-----FGVGTFRYNFNPFFYAKKGLGIK-----TKYDV-- 126
 Db 111 SMSGGYWRDLDLNEVSVNTLMNLNAYDFRNDSAFTYWSAG--IGAKELHOKTTGISTWD 169
 QY 126 -----TSRNATYYSNKSDKTSLAGVGFGFKPLANVGVEASNYLSEDA 173
 Db 170 YGSGSGRSLSSGSDADNFAMSLAGVRYDVTPIALDLSTRYLDAGSSVS 222

RESULT 9

Q46678 PRELIMINARY: PRT: 264 AA.

AC 046678: PRELIMINARY: PRT: 264 AA.
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE HEAT RESISTANT AGGLUTININ 1 PRECURSOR.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O9:H10:K99;
 RX MEDLINE: 95012721.
 RA LUTWYCHE P., RUPPS R., CAVANAGH J., WARREN R.A., BROOKS D.E.;
 RT "Cloning, sequencing, and viscometric adhesion analysis of heat-
 resistant agglutinin 1, an integral membrane hemagglutinin from

RT Escherichia coli O9:H10:K99."
 RL Infect. Immun. 62:5020-5026(1994).
 DR EMBL: U07174; AAC13751.1; -;
 DR PFAM: PF01389; OmpA_membrane; 1.
 KW Signal.
 FT SIGNAL 1
 FT CHAIN 40
 SQ SEQUENCE 264 AA; 29024 MW; 77A084A2 CRC32;
 POTENTIAL.
 HEAT RESISTANT AGGLUTININ 1.

Query Match 12.0%; Score 109.5; DB 2; Length 264;
 Best Local Similarity 21.0%; Pred. No. 0.12; Mismatches 75; Indels 71; Gaps 10;

Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;
 QY 1 MKTLKTLAVSASSLAMSANAIS-----XGNSADAQP 34
 Db 15 MEMKNVIAVSLAMAGMSTQALADESKTGFYVGKAGASVSLADQRFSLNGEETSK 74
 QY 35 YVAKTIGVDAKOINGKNAY--GIYAGYNPDONGV---EAEPV--GSDAKEFNAGVS 86
 Db 75 YKGG-----DGHDFVFSGIAAGYDFYPOFSIPVTELEFYARGLKADSKYNVDKD 124
 QY 87 PVKG-----DVKS-----FGVGTFRYNFNPFFYAKKGLGIK-----TKYDV-- 126
 Db 125 SMSGGYWRDLDLNEVSVNTLMNLNAYDFRNDSAFTYWSAG--IGAKELHOKTTGISTWD 183
 QY 126 -----TSRNATYYSNKSDKTSLAGVGFGFKPLANVGVEASNYLSEDA 173
 Db 184 YGSGSGRSLSSGSDADNFAMSLAGVRYDVTPIALDLSTRYLDAGSSVS 236

RESULT 10

Q33796 PRELIMINARY: PRT: 165 AA.

AC 033796: PRELIMINARY: PRT: 165 AA.
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
 DE ATTACHMENT AND INVASION PROTEIN HOMOLOG.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.

RN [1]
 RP SEQUENCE FROM N.A.
 RA DE GROOTE M., OCHSNER U.A., SHILOH M., NATHAN C., MCCORD J.M.,
 RA DINARD M.C., LIBBY S.J., VAZQUEZ-TORRES A., XU Y., FANG F.C.;
 RL Submitted (JUN-1997) to the EMBL/genbank/DBJ databases.
 DR EMBL: AF007380; AAB62386.1; -;
 SQ SEQUENCE 165 AA; 17439 MW; C70DICB3 CRC32;

Query Match 12.0%; Score 109; DB 2; Length 165;
 Best Local Similarity 25.6%; Pred. No. 0.07; Mismatches 68; Indels 32; Gaps 6;

Matches 41; Conservative 19; Mismatches 68; Indels 32; Gaps 6;
 QY 36 VGAKTIGVDAKOING-KNT-----AYGIYAGYNF---DONGVEAEFV 74
 Db 10 VGLALGSGVNRNAGYKKNVYSIGYATDLSGMLSGANANGANKYNNEDDLSGGMAGSVT 69
 QY 75 GSDAKEFNAGVPGKDVKSFGAYGTYRIN-FINPFFYAKKGLIATKYVDVTSRNATYV 133
 Db 70 YNSADVNNNGYKKGDADYSLVGPSPYRNDYLN---AYVMIGANGHI-----KDNW 119
 QY 134 SNKSDKTSLAGVGFGFKPLANVGVEASNYLSEDA 173
 Db 120 GNSDNKTAFAFGAGIQLNPVENIAVNASYEHTSFTDADS 159

RESULT 11

Q46680 PRELIMINARY: PRT: 247 AA.

AC 046680: PRELIMINARY: PRT: 247 AA.
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

RP SEQUENCE FROM N.A.
 RC STRAIN-F38;
 RX MEDLINE: 98453456.
 RA ITO Y., AZUMA T., SUTO H., MIYAJI H., YAMAZAKI Y., KOHLI Y.,
 RA KURIYAMA M.;
 RT "Full-length sequence analysis of the vacA gene from cytotoxic and
 noncytotoxic Helicobacter pylori."
 RL J. Infect. Dis. 178:1391-1398(1998).
 DR EMBL: AF049643; AAD04279.1; -
 SQ SEQUENCE 1291 AA; 139427 MW; 9E3761BB CRC32;

Query Match 11.4%; Score 104; DB 2; Length 1291;

Best Local Similarity 25.4%; Pred. No. 2.5;
 Matches 54; Conservative 28; Mismatches 75; Indels 56; Gaps 12;

QY 9 AVSASSILMSANAAISYGSNADAPYVGAKI-----GQVDAKQINGKNTAYGIYAGY 61
 Db 1029 AIGGASLNS-GGNTSL-YGTSAGVDALNKEVLAIVGFGSYGSSFN--NQANSLNSGA 1084
 QY 62 NFDQNGV-----EAEF-----VGSDAKEFNAGVSPKGVKSGF--AYGT----- 101
 Db 1085 N-NANFGVYSRIIPANRHEFDEFOAGAVGSDQSLNFKSALLRDLNOSYNVLAYGAATRAS 1143
 QY 101 --YRYNFINTPFYAKGKLGIAKTKVDYTSRNATYTSNKSDDKSLAGCVGVGFKPLANVGY 158
 Db 1144 YGYDFAFFRNALVLPKPSGVSYNHLGSTNFE--SNSTKTKALKNGASSOHLFNASANY 1199
 QY 159 EASYNY-----LSE-----DANAISL 174
 Db 1200 EARYYYGDTSYFYMNAGVLOEFANFGSSNNAVSL 1232

RESULT 15

Q9ZHT8 PRELIMINARY; PRT; 1291 AA.

AC 09ZHT8;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, last annotation update)
 DE VACUOLATING CYTOTOXIN PRECURSOR.
 GN VACA.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-F64;
 RX MEDLINE: 98453456.
 RA ITO Y., AZUMA T., ITO S., SUTO H., MIYAJI H., YAMAZAKI Y., KOHLI Y.,
 RA KURIYAMA M.;
 RT "Full-length sequence analysis of the vacA gene from cytotoxic and
 noncytotoxic Helicobacter pylori."
 RL J. Infect. Dis. 178:1391-1398(1998).
 DR EMBL: AF049647; AAD04282.1; -
 SQ SEQUENCE 1291 AA; 139395 MW; 2B366B27 CRC32;

Query Match 11.4%; Score 104; DB 2; Length 1291;

Best Local Similarity 25.4%; Pred. No. 2.5;
 Matches 54; Conservative 28; Mismatches 75; Indels 56; Gaps 12;

QY 9 AVSASSILMSANAAISYGSNADAPYVGAKI-----GQVDAKQINGKNTAYGIYAGY 61
 Db 1029 AIGGASLNS-GGNTSL-YGTSAGVDALNKEVLAIVGFGSYGSSFN--NQANSLNSGA 1084
 QY 62 NFDQNGV-----EAEF-----VGSDAKEFNAGVSPKGVKSGF--AYGT----- 101
 Db 1085 N-NANFGVYSRIIPANRHEFDEFOAGAVGSDQSLNFKSALLRDLNOSYNVLAYGAATRAS 1143
 QY 101 --YRYNFINTPFYAKGKLGIAKTKVDYTSRNATYTSNKSDDKSLAGCVGVGFKPLANVGY 158
 Db 1144 YGYDFAFFRNALVLPKPSGVSYNHLGSTNFE--SNSTKTKALKNGASSOHLFNASANY 1199

QY 159 EASYNY-----LSE-----DANAISL 174
 Db 1200 EARYYYGDTSYFYMNAGVLOEFANFGSSNNAVSL 1232

Search completed: May 20, 2000, 12:22:04
 Job time: 4021 sec

THIS PAGE BLANK (USPTO)